In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institute shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the Dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

7/25/68

Original Page Numbering Retained.

PERIODIC ORBITS OF A DYNAMICAL SYSTEM

A THESIS

Presented to

The Faculty of the Division of Graduate

Studies and Research

by

Philip F. Lee, III

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in the School of Mathematics

Georgia Institute of Technology

July, 1970

PERIODIC ORBITS OF A DYNAMICAL SYSTEM

Approved:

Chairman

Date approved by Chairman July 31, 1970

ERRATUM

Page 59 does not appear in the following text; this is an error in numbering only; none of the text is lacking.

ACKNOWLEDGMENTS

I wish to thank Dr. Kurth for his efforts in the fruition of this work. I am grateful also for many hours of discussions and lectures during which he tried to develop more than just a technical facility with mathematics. Knowing what I do now, I wish that it were possible to re-run them all. Thanks are due to Drs. Sledd and Ford for reading the thesis — I am especially grateful to Dr. Sledd for his criticism of my work and his help in improving it. I am pleased to acknowledge a debt of long standing to Dr. Osborn for guidance which spans my undergraduate and my graduate residence. The assistance mentioned above in my flying lessons implies no shared responsibility for any crash landings that I might make.

I am grateful to the taxpayers of the State of Georgia and of the United States who through their agencies (the University System of Georgia and the National Aeronautics and Space Administration) financially supported my education; my gratitude is not diminished by the arrogant bureaucrats who administer the funds here.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

1.1. Dynamical Systems

In 1902 E. T. Whittaker published a theorem which asserted the existence of periodic orbits of certain dynamical systems of two degrees of freedom (see [Wh; pp. 386-389]). We shall generalize Whittaker's theorem and prove our generalization. In order to formulate the theorem precisely, we immediately introduce and discuss some concepts.

A particle or point mass is an ordered pair $(\mu, \underline{x})$, where μ is a positive real number and $\underline{x}$ is an n-tuple of real numbers — $\underline{x} = (x^1, x^2, \ldots, x^n)$. The positive number μ is called the mass of the particle and the n-tuple $\underline{x}$ is called its position.

A motion of m particles in Euclidean n-space E^n is a finite sequence of pairs

$$(\mu_1, \underline{f}_1(t)), \ldots, (\mu_m, \underline{f}_m(t)),$$

where $\underline{f}_i(t)$ is a continuous function with domain the non-empty open interval (a,b) of the real line (time) and range a subset of Euclidean n-space for each integer $i = 1, \ldots, m$.

1.1.1. Newtonian Dynamical Systems; Conservation of Energy

A motion of m particles in Euclidean n-space is called a (conservative) Newtonian dynamical system if each function $\underline{f}_i$ has at least two continuous derivatives on the non-empty open interval (a,b) of the "time" axis and on this interval satisfies a differential equation of the form

$$\mu_i \underline{\ddot{f}}_i(t) = - V_{\underline{x}_i}(\underline{f}_1(t), \ldots, \underline{f}_m(t)), \quad i = 1, \ldots, m. \tag{1}$$

The dots over the functions $\underline{f}$ in equations (1) indicate second derivatives with respect to the argument of the functions; the letter V denotes a real-valued function, defined and continuous and having continuous first partial derivatives on an open subset of E^{nm} (Euclidean nm-space); $V_{\underline{x}_i}$ denotes the n-tuple of functions

$$\left(\frac{\partial V}{\partial x_i^1}, \ldots, \frac{\partial V}{\partial x_i^n} \right) .$$

The function V is called the potential of the Newtonian dynamical system.

The next proposition is usually known by the title of conservation of energy.

Proposition 1. For the Newtonian dynamical system $(\mu_1, \underline{f}_1(t))$, $\ldots$, $(\mu_m, \underline{f}_m(t))$ in Euclidean n-space with potential V, the expression

$$\tfrac{1}{2} \sum_{i=1}^{m} \mu_i \sum_{j=1}^{n} [\dot{f}_i^j(t)]^2 + V(\underline{f}_1(t), \ldots, \underline{f}_m(t)) \tag{2}$$

is a constant function of the time variable on the interval (a,b).

We denote the constant value of the expression (2) by the letter h. The real number h is called the energy of the dynamical system $\left\{ (\mu_i, \underline{f}_i(t)) \right\}_{i=1}^{m}$.

This proposition is proved by differentiating expression (2) with respect to the time t and applying equation (1) to show that the derivative is zero.

1.1.2. Maupertuis' Principle (Least Action)

The two propositions of this subsection use concepts and results from the calculus of variations. For the relevant material see Appendix A.

Proposition 1. Suppose that the m-tuple of functions $(\underline{\phi}_1(s), \ldots, \underline{\phi}_m(s))$ satisfies the condition

$$h - V(\underline{\phi}_1(s), \ldots, \underline{\phi}_m(s)) > 0$$

for each real number s in the interval $[0, s_1]$, and suppose that this m-tuple yields a stationary value of the integral

$$\int_0^{s_1} \left\{h - V(\underline{x}_1(s), \ldots, \underline{x}_m(s))\right\}^{\frac{1}{2}} \left\{\tfrac{1}{2}\sum_{i=1}^{m} \mu_i \sum_{j=1}^{n} [\dot{x}_i^j(s)]^2\right\}^{\frac{1}{2}} ds, \tag{3}$$

where the parameter s is selected so that

$$\left\{h - V(\underline{\phi}_1(s), \ldots, \underline{\phi}_m(s))\right\} \left\{\tfrac{1}{2}\sum_{i=1}^{m} \mu_i \sum_{j=1}^{n} [\dot{\phi}_i^j(s)]^2\right\} = 1 . \tag{4}$$

Let the real-valued function s(t) be defined by the equation

$$\frac{ds}{dt} = h - V(\underline{\phi}_1(s), \ldots, \underline{\phi}_m(s))$$

and the initial condition $s(0) = 0$. Let the functions $\underline{\Psi}_1(t), \ldots, \underline{\Psi}_m(t)$ be defined by

$$\underline{\Psi}_i(t) = \underline{\phi}_i(s(t)), \quad i = 1, \ldots, m.$$

Then the motion $\left\{(\mu_i, \underline{\Psi}_i(t))\right\}_{i=1}^{m}$ is a Newtonian dynamical system which is defined for times in the interval $[0,t_1]$ (t_1 is defined by the equation $s(t_1) = s_1$); the system has potential V and energy h.

<u>Proposition 2</u>. Conversely, let $\left\{(\mu_i, \underline{\Psi}_i(t))\right\}_{i=1}^{m}$ be a Newtonian dynamical system on an open interval containing the interval $[0,t_1]$. Denote the potential of the system by the letter V and the energy of the system by the letter h. Define n-tuples $\overset{o}{\underline{x}}_i$ and $\overset{o}{\underline{v}}_i$ by the equations

$$\underline{\Psi}_i(0) = \overset{o}{\underline{x}}_i$$

$$\frac{d\underline{\Psi}_i}{dt}(0) = \overset{o}{\underline{v}}_i \, [h - V(\overset{o}{\underline{x}}_1, \ldots, \overset{o}{\underline{x}}_m)] ,$$

$i = 1, \ldots, m$; define the function $t(s)$ by the equation

$$\frac{dt}{ds} = \frac{1}{h - V[\underline{\Psi}_1(t), \ldots, \underline{\Psi}_m(t)]}$$

and the initial condition $t(0) = 0$. Suppose that for t in the interval $[0, t_1]$

$$h - V\,[\underline{\Psi}_1(t), \ldots, \underline{\Psi}_m(t)\,] > 0.$$

Let the function $\underline{\phi}_i(s)$ be defined by the equation

$$\underline{\phi}_i(s) = \underline{\Psi}_i(t(s)), \quad i = 1, \ldots, m.$$

Then $\left\{\underline{\phi}_i(s)\right\}_{i=1}^{m}$ yields a stationary value of the integral (3) while satisfying equation (4).

These two propositions taken together are the Jacobi formulation of the Maupertuis principle (see [Wi; pp. 122-127 and 417]). A proof of them can be constructed by comparing the Euler equation (see Appendix A) of the integrand in (3) with Newton's equation (1). We defer proving these propositions until Section 2.6, below, where they are given in a more general setting.

The integral (3) in the previous subsection has the same form as length of arc of a curve lying on a surface in Euclidean N-space for some integer $N \geqq mn + 1$. The study of such forms and the sets on which they are defined falls within the area of Riemann geometry. Needed results from Riemann geometry will be developed in Chapter Two. In the last part of Chapter Two we shall draw more precisely the connection indicated above between Newtonian dynamical systems and Riemann geometry.

1.2. Whittaker's Theorem and the Problem

Whittaker's theorem (see [Wh; pp. 386-389]) asserts the existence of periodic orbits of a Newtonian dynamical system of two degrees of freedom. Such a system is described by two real functions f_o and g_o of the real (time) variable t. These functions are defined and have continuous derivatives through second order on an open (non-empty) interval I_o of the real line. It is assumed that f_o and g_o satisfy Newton's equations of motion on I_o, i.e.,

$$\ddot{f}_o(t) = - V_1(f_o(t), g_o(t)),$$

$$\ddot{g}_o(t) = - V_2(f_o(t), g_o(t)).$$

Here V denotes the given potential per unit mass of the system; V_1 and V_2 denote the partial derivatives of this potential with respect to its first and second variables respectively and the double dots over f_o and g_o denote second derivatives with respect to t. The function V and its partial derivatives are assumed to be defined and continuous in an open (non-empty) subset of the Euclidean (x,y)-plane.

Whittaker considers a closed subset of the Euclidean plane which is topologically equivalent to a closed annulus. By this statement we mean that there is a one-to-one function with domain the annulus and range the closed subset and this function is continuous and has a continuous inverse. The closed subset is bounded by two Jordan curves. By a Jordan curve we understand the range of an ordered pair of real-valued continuous functions (f,g) defined on a closed interval $[a,b]$ of the real line, where the functions f and g are such that if $f(t_1) = f(t_2)$ and $g(t_1) = g(t_2)$ for $t_1<t_2$, then $t_1 = a$ and $t_2 = b$. Whittaker further supposes that the periodic extentions of f and g have continuous derivatives through second order.

According to the Jordan curve theorem each of the bounding curves separates the plane into a bounded portion and an unbounded portion. By the outer normal of a Jordan curve we understand the perpendicular to the tangent, the positive sense of the perpendicular being toward the unbounded portion of the plane. Whittaker denotes by γ the oriented angle between the x-axis and the outer normal of either of the boundary curves and by ρ the radius of curvature of either of the curves at an arbitrary point.

Whittaker's assertion reads [Wh; p. 389]:

> If one closed curve be enclosed by another closed curve, and if the quantity
>
> $$\frac{h - V(x,y)}{\rho} - \tfrac{1}{2}(\cos\gamma)\frac{\partial V}{\partial x} - \tfrac{1}{2}(\sin\gamma)\frac{\partial V}{\partial y} \tag{1}$$
>
> be negative at all points of the inner curve and positive at all points of the outer curve, then in the ring-shaped

space between the two curves there exists a periodic orbit of the dynamical system for which the constant of energy is h.

The term "periodic orbit" means a pair of functions f_o and g_o satisfying Newton's equations of motion (above) for all time t and having the same positive period T. That is,

$$f_o(t + T) = f_o(t)$$

and

$$g_o(t + T) = g_o(t)$$

for all t. The energy constant h of the dynamical system (f_o, g_o) is the constant value of

$$\tfrac{1}{2}\left\{ (\dot{f}_o(t))^2 + (\dot{g}_o(t))^2 \right\} + V(f_o(t), g_o(t)).$$

The quantity

$$\frac{h - V(x,y)}{\rho} - \tfrac{1}{2}(\cos \gamma)\frac{\partial V}{\partial x} - \tfrac{1}{2}(\sin\gamma)\frac{\partial V}{\partial y}$$

arises from the Jacobi formulation of the Maupertuis principle (see Section 1.1). Jacobi has shown that Newton's equations of motion are equivalent to the formulation of the Maupertuis principle stated in Subsection 1.1.2. In the present case (two degrees of freedom) by this equivalence the functions f_o and g_o yield a stationary value of the integral

$$I(f,g) = \int_a^b \left\{h - V(f(t),g(t))\right\}^{\frac{1}{2}} \left\{[\dot{f}(t)]^2 + [\dot{g}(t)]^2\right\}^{\frac{1}{2}} \quad dt.$$

Whittaker's argument consists essentially of the following four steps:

1. The condition that

$$"\frac{h - V(x,y)}{\rho} - \tfrac{1}{2}\cos\gamma\,\frac{\partial V}{\partial x} - \tfrac{1}{2}\sin\gamma\,\frac{\partial V}{\partial y}$$

   be negative at all points of the inner curve and positive at all points of the outer curve" is shown to imply that any closed curve which is sufficiently close to either of the boundary curves has I-value less than the I-value of the boundary curve.

2. Whittaker seeks a closed curve having a minimum I-value since such a closed curve also yields a stationary value for I.

3. Because of step one, Whittaker asserts that a closed curve yielding a minimum I-value exists in the interior of the ring-shaped region.

4. From the equivalence of the Maupertuis principle to Newton's equations of motion, this curve is a periodic orbit of the dynamical system.

In this argument a gap occurs in step three. Whittaker has not demonstrated the existence of the curve. He has only shown that if it exists, it cannot intersect the boundary curves. One of the purposes of this investigation is to establish the existence of such a curve.

1.3. Observations of a Physical and Geometrical Nature

In his argument Whittaker seeks a closed curve which minimizes the integral I (see the previous section). The expression

$$\left\{h - V(x,y)\right\}^{\frac{1}{2}} \left\{dx^2 + dy^2\right\}^{\frac{1}{2}}$$

may be regarded as the first fundamental form defining length of arc on a Riemann manifold [Wi; p.127]. A closed curve which yields a minimum for I becomes a closed curve which has minimum length.

To continue our intuitive discussion we need the well-known fact (see [Wd; pp. 33-39] or Section 2.5 below) that under appropriate hypotheses any two neighboring points of a Riemann manifold may be connected by a "unique" curve having length smaller than any other curve joining the two points.

The expression (1) of Section 1.2 (see Whittaker's theorem) may be interpreted physically as one-half the difference between the centrifugal force for a fictitious motion of energy h along the boundary,

$$\frac{h - V(x,y)}{\rho} = \frac{\frac{1}{2}mv^2}{\rho} ,$$

and the component of the force due to the potential V in the direction of the inner normal. Since for a dynamical system the difference is zero and since for the outer boundary curve the difference is assumed positive, at the point of tangency the radius of curvature for a dynamical system tangent to the outer curve is greater than the radius of curvature of the boundary curve. Because of this difference in radii of curvature, the tangent dynamical system cannot enter the ring-shaped region for values

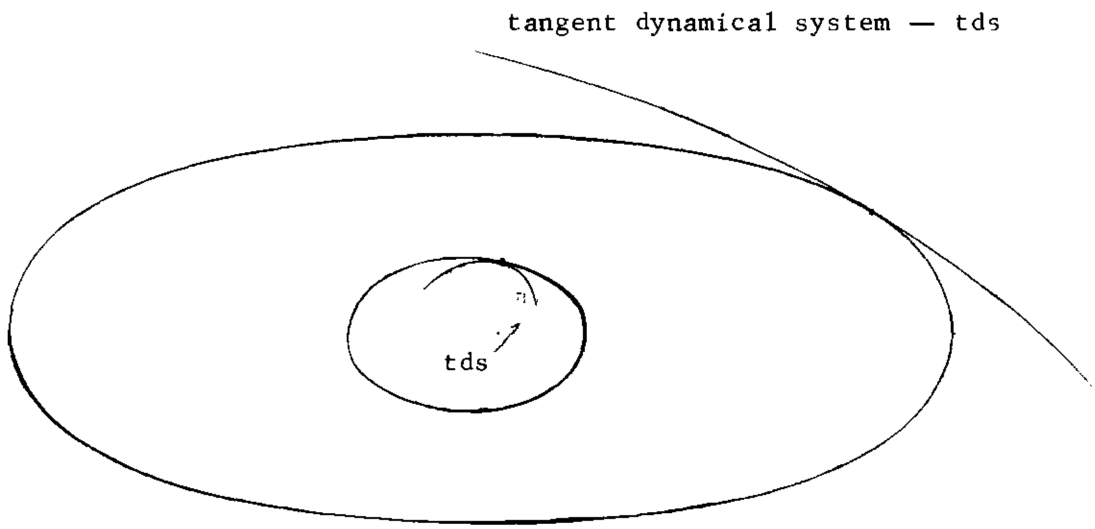

Figure 1. Behavior of dynamical systems which are tangent to the boundary curves of Whittaker's ring-shaped region.

of the time near the time of contact with the boundary curve (see Figure 1). This fact is established in Chapter Four below. It can be established here by introducing a system of perpendicular coordinates at the point of tangency with one axis tangent to the outer curve. The functions describing the boundary curve and the tangent dynamical system are expanded by Taylor's theorem and the difference in second order terms yields the result.

A similar conclusion may be drawn for the inner boundary curve. As we shall see, the result of these conclusions is that the unique shortest curve (dynamical system) which joins two neighboring points on the boundary of the ring-shaped region lies in the interior of the region except for its "endpoints." This convexity observation provides the basis for our development in Chapters Two and Three.

We approach the proof of Whittaker's theorem from a geometrical viewpoint with straight lines of Euclidean geometry replaced by the aforementioned shortest curves. We shall show that the hypotheses of Whittaker's theorem insure that the ring-shaped region is locally convex. A locally convex set, we understand, is a set such that any two of its points which are sufficiently close together may be joined by a curve of least length lying in the set. The notion of local convexity is used to generalize the Whittaker boundary hypotheses to cover dynamical systems of n degrees of freedom in the guise of a Riemann manifold of dimension n (see Chapter Four below).

In the process of proving Whittaker's theorem, we establish a generalization of a theorem of Tietze [T; p. 697] , who showed that a closed locally convex subset of n-dimensional Euclidean space is convex.

A side result of our arguments will be that (under suitable hypotheses) a locally convex set is also "convex" (any two points can be joined by a shortest curve which lies in the set).

1.4. Some Historical and Other Considerations

In [B] G. D. Birkhoff attributes the first rigorous proof of Whittaker's theorem to Signorini [S] and claims to extend Whittaker's and Signorini's results. Birkhoff restricts his attention to dynamical systems of two degrees of freedom, and he relaxes the conditions on the ring-shaped region [B; pp. 216-219] to the point that the boundary need not be given by two parametric curves. As Birkhoff points out, the nature of his generalization implies that his region is a "limit" of an increasing sequence of ring-shaped regions whose boundaries are given by parametric curves which satisfy the convexity conditions of Section 1.2 [B; p.217]. Birkhoff's article is difficult to read and lacking in rigor.

We shall prove Whittaker's theorem by using techniques first developed by Hilbert [Bz; pp. 419-443]. Because of the many developments in the geometry of metric spaces (see works of Blumenthal, Busemann, Menger and Rinow), it is possible to give intuitively pleasing geometrical arguments. We generalize the theorem to apply to dynamical systems of n degrees of freedom. It is for this generalization that the concept of Riemann manifold becomes important.

In order to state the generalization, the next chapter is devoted to developing the necessary tools from Riemann geometry.

CHAPTER II

RIEMANN MANIFOLDS AND DYNAMICAL SYSTEMS

In Chapter One a connection was indicated between Whittaker's theorem and Riemann manifolds (Maupertuis' principle — see Sections 1.1 and 1.2). In this chapter we shall carefully draw this connection in rigorous terms. The term "Riemann manifold," already used in a vague sense, is defined and certain properties of such manifolds are investigated. The investigation of Riemann manifolds centers around the properties of convexity which we observed to hold for the ring-shaped region occuring in the Whittaker theorem (see Section 1.3).

The notion of a Riemann manifold provides the means of generalizing Whittaker's theorem. However, the details of Riemann geometry obscure the basic ideas involved in our arguments. For this reason once we have developed the basic results of Riemann geometry, we use these results as axioms for a class of "distance spaces." It is in this setting that Whittaker's theorem is proved, and in this setting the geometrical nature of our arguments becomes more perspicuous.

We begin this chapter with some basic notions from topology. Manifolds, tensors and Riemann manifolds are discussed in succeeding sections. Since our mechanical problem and the existence of shortest curves in a Riemann manifold give rise to ordinary differential equations, we recall an existence and uniqueness theorem concerning solutions of such equations. In the last part of Chapter Two the actual convexity properties of Riemann manifolds are discussed, and the connection between Riemann manifolds and dynamical systems is given.

2.1. Topology

In this section definitions and results from point-set topology are given in a form which proves useful in our development.

2.1.1. Topological Spaces

Let T be a collection of sets such that any union of members of T and any finite intersection of members of T belong again to the set T. If the empty set belongs to T, then we say that the set T is a topology on the space X where the set X is defined by

$$X = \bigcup \{ G \mid G \text{ is a member of } T \}.$$

The set X is called the space of the topology T. We shall use the (logically incorrect) terminology "topological space X" to label the understood topology T as well as its space X. The members of the topology T are called the open sets (of X). The complement (in X) of an open set is called a closed set (of X).

Let A denote a subset of the topological space X. A point x in the set X is called a limit point of the set A if every open set containing x has a non-empty intersection with the set $A - \{x\}$. The closure of a set A in the space X, denoted by cl(A), is defined to be the union of the set A with the set of its limit points. Clearly, the closure of a set A is a closed set. If A is closed, then A and its closure are identical.

A point x is called an interior point of a set A if there is an open set containing the point x and this open set is a subset of A. The set of all interior points of A, denoted by Int(A), is called the interior of A. Clearly, the interior of a set is an open (possibly empty) set. If the set A is open, then Int(A) and A are identical. An open set containing a point x will be called a neighborhood of x.

2.1.2. Distance Function

A non-negative, real-valued function d which is defined on the cartesian product X x X is called a distance function on the set X if the following conditions are satisfied:

1. the number $d(x,y)$ vanishes if and only if x is identical with y;
2. $d(x,y) = d(y,x)$ for all points x and y in X; and
3. $d(x,z) \leqq d(x,y) + d(y,z)$ for all points x, y and z in X (the triangle inequality).

If d is a distance function on the set X, we say that the set X is a distance space. The set of points

$$S(x,\epsilon) = \{y \epsilon X \mid d(x,y) < \epsilon\}$$

is called the ϵ-ball about x. A set G in a distance space X is said to be open if, for any point x in G, there is a positive number ϵ such that the ϵ-ball $S(x,\epsilon)$ is a subset of G. Let T_d denote the collection of all open sets of a distance space, then T_d is a topology on the space X. The topology T_d is called the topology of the distance function d. Such a topology will be called a distance topology. The triangle inequality implies that ϵ-balls are open sets for each positive number ϵ.

One important property of distance topologies is that any two distinct points x and z of X have disjoint neighborhoods. In fact let

$r = d(x,z) > 0$. An application of the triangle inequality shows that the balls

$$S(x,r/2) \qquad \text{and} \qquad S(z,r/2)$$

are disjoint. A topology not arising from a distance function may or may not have this property. The term Hausdorff space, we understand, refers to a topological space in which any two distinct points have disjoint neighborhoods.

2.1.3. Mappings between Topological Spaces

Let f denote a function or mapping with domain the subset A of the topological space X. Suppose that the range of the function f is a subset of the topological space Y. The function f is called continuous on A (if A = X, just continuous) if the preimage of any open set of Y is the intersection of an open set in X with the set A. The preimage of a subset B of Y is the set

$$f^{-1}(B) = \left\{ x \mid f(x) \varepsilon B \right\} .$$

Let A and B be subsets of the topological spaces X and Y respectively. Suppose that the function f is continuous and one-to-one with domain A and range B. Suppose also that the function inverse to f, f^{-1}, is continuous. Then the sets A and B are called topologically equivalent. Topological equivalence is an equivalence relation. If A and B are topologically equivalent, we say also that A and B are homeomorphic. The mapping f which is one-to-one and continuous in both directions is called a homeomorphism.

If the domain A of the continuous function f is a closed interval of the real line, then we call this function a parametric curve. In the case of a parametric curve it is evident that the image of A under the function f cannot be written in the form

$$(f(A)\cap G_1)\cup(f(A)\cap G_2),$$

where G_1 and G_2 are open, non-empty sets.

A topological space which is not the union of two open, non-empty sets is called <u>connected</u>. A subset B of a topological space is called connected if

$$B \neq (B\cap G_1)\cup(B\cap G_2)$$

for any two open, non-empty sets of G_1 and G_2.

In this terminology any interval of the real line is connected. Further, it is evident that the continuous image of a connected set is connected. We shall restrict our attention to connected topological spaces.

2.1.4. <u>Sequences</u>

A function whose domain is some subset of the non-negative integers is called a <u>sequence</u>. A sequence is denoted by $\{x_n\}_{n=0}^{\infty}$ or $\{x_n\}_{n\in S}$ where in the first case the domain is the set of non-negative integers and in the second case the domain is the set S. The value of the sequence $\{x_n\}_{n\in S}$ at the integer n is denoted by x_n. If $\{x_n\}_{n\in S}$ is a sequence and if W is a subset of S, then the sequence $\{x_n\}_{n\in W}$ is called a <u>subsequence</u> of the sequence $\{x_n\}_{n\in S}$. A sequence whose domain is a finite set is called <u>finite</u>; otherwise, a sequence is called <u>infinite</u>.

Let X be a topological space. Suppose that there is a sequence $\{G_n\}_{n=0}^{\infty}$ in which each G_n is an open set, and suppose that any open set U can be written as a union of the range of a subsequence of the sequence $\{G_n\}_{n=0}^{\infty}$. In this case we say that the topological space X has a <u>countable base</u>.

Let X be a topological space, and let $\{x_n\}_{n\in S}$ be an infinite sequence of points in X. We say that this sequence has <u>limit x</u>, and we write symbolically

$$x_n \xrightarrow[S]{} x$$

if any neighborhood of the point x contains all of the points x_n except (possibly) for finitely many integers n. We also say that the sequence $\{x_n\}_{n\in S}$ <u>converges</u> to the point x. In view of the definition of the term Hausdorff space, it is evident that in such a space a sequence can have at most one limit. In the case that the sequence $\{x_n\}_{n\in S}$ has the unique limit x, we write symbolically

$$x = \lim x_n \ (n\in S)$$

in place of

$$x_n \xrightarrow[S]{} x.$$

Let f be a continuous function with domain the set A (A is a subset of the topological space X) and with range a subset of the topological space Y. Let $\{x_n\}_{n\in S}$ be a sequence of points from the set A. Suppose that $x = \lim x_n \ (n\in S)$ is a point in the set A. Then
$f(x) = \lim f(x_n) \ (n\in S)$.

2.2 Manifolds and Their Tangent Spaces

We assume in this section that all topological spaces are Hausdorff and connected and have a countable base.

2.2.1. Local Coordinates and Atlases

A topological space X is called a topological manifold of dimension n if each point has a neighborhood which is topologically equivalent to n-dimensional Euclidean space E^n. Since for different integers n and m the Euclidean spaces E^n and E^m are not topologically equivalent [HW; pp. 24 and 41], the dimension of a topological manifold is uniquely determined.

Let U be an open subset of X which is topologically equivalent to E^n. Let f be a homeomorphism between U and E^n (or a subset of E^n). The pair (U,f) is called a chart or a system of local coordinates in X. If $x^1, \ldots, x^n$ are coordinates in E^n of the point f(x) (x a point of U), then $x^1, \ldots, x^n$ are called local coordinates of the point x.

By the term atlas of class C^p (p a positive integer) on the topological manifold X we mean a sequence of charts $\{(U_i,f_i)\}_{i\in S}$ with the following properties:

1. each point of X is contained in at least one set U_i, and
2. if the intersection of the sets U_i and U_j is not empty, then the composite topological mappings $f_j f_i^{-1}$ and $f_i f_j^{-1}$ have continuous partial derivatives of all orders through p.

Since the mapping $f_j f_i^{-1}$ is inverse to the mapping $f_i f_j^{-1}$ and since both mappings have continuous partial derivatives, their Jacobian determinants cannot vanish.

Two atlases of class C^p are said to be equivalent if their union is an atlas of class C^p. This definition of equivalence yields an equivalence relation [A; pp. 6-7]. A differentiable structure of class C^p is said to be defined on X if on X there is defined an atlas of class C^p. In the case that X has a differentiable structure of class C^p, we say that X is a differentiable manifold of class C^p.

Let X be a differentiable manifold of class C^p with differentiable structure given by the atlas $\{(U_i, f_i)\}_{i \in S}$. The chart (U,f) is said to be allowable if the atlas $\{(U_i, f_i)\}_{i \in S} \cup \{(U,f)\}$ is equivalent to the atlas $\{(U_i, f_i)\}_{i \in S}$.

2.2.2. Real-Valued Functions on Differentiable Manifolds; Parametric Curves

Let X denote an n-dimensional differentiable manifold of class C^p. Let q be a positive integer not exceeding the integer p and suppose that α is a real-valued function with domain a subset of X. Let x be a point of the domain of α and let (U,f) be an allowable chart with U containing x. If the function αf^{-1} is continuously differentiable of class C^q at the point f(x) in E^n, then we say that the function α is continuously differentiable of class C^q at the point x. If the function α is continuously differentiable of class C^q at each point of X, then we say that α is of class C^q on X and we write symbolically $\alpha \in C^q(X)$. This definition is independent of which atlas in an equivalence class of atlases is used to define the differentiable structure of X.

Let w denote a parametric curve with domain the interval [a,b] and range a subset of the differentiable manifold X of class C^p. Let t_o be a real number in the open interval (a,b) and let (U,f) be an allowable chart with U containing the point $w(t_o)$. If the parametric curve $\underline{w} = fw$ in Euclidean n-space is continuously differentiable on an open interval containing t_o, then we say that the parametric curve w is continuously differentiable at the point t_o. If the parametric curve w is continuously differentiable at each point of the open interval (a,b), then we say that the curve w is continuously differentiable on the interval (a,b) and write symbolically $w \epsilon C^1(a,b)$. The parametric curve w is said to be continuously differentiable on the closed interval $[c,d]$ if it is continuously differentiable on an open interval containing the interval $[c,d]$.

Let $w^i = f^i w$ be the ith coordinate of the continuously differentiable parametric curve w. Let $\underline{w}$ denote the n-tuple $(w^1, \ldots, w^n)$. Then w is continuously differentiable (by definition) if and only if $\underline{w}$ is continuously differentiable. At each t_o in the interval (a,b), the n-tuple $\dot{\underline{w}}(t_o)$,

(Def) $$\dot{\underline{w}}(t_o) = \frac{d}{dt}\left\{\underline{w}(t)\right\}_{t=t_o} ,$$

is interpreted as coordinates of the tangent vector to the parametric curve $\underline{w}$ in E^n. In the next few pages we define the tangent vector to the curve w in the differentiable manifold X and relate it to the n-tuple $\dot{\underline{w}}$.

2.2.3. Tangent Spaces, Tangent Vectors and Their Coordinates

Let χ_w be the linear function defined as follows:

(Def) $$\chi_w(\alpha) = \frac{d}{dt}\left\{\alpha w(t)\right\}_{t=0}$$

where α is any real-valued function which is continuously differentiable at the point x_o in X, where w is a parametric curve which is continuously differentiable at $t = 0$, and where $w(o) = x_o$. The parametric curves ω and w are said to have the <u>same tangent vector (at $t = 0$)</u> if $\chi_\omega(\alpha)=\chi_w(\alpha)$ for all real-valued functions α which are continuously differentiable at the point $x_o = \omega(o) = w(o)$. "Having the same tangent vector" is an equivalence relation among all parametric curves w which are continuously differentiable at $t = 0$ with $w(0) = x_o$. An equivalence <u>class</u> of such parametric curves will be called a <u>tangent vector</u> of the manifold X at the point x_o.

We next define arithmetic operations on tangent vectors. We also show that the resulting vector space has dimension n. This vector space will be called the tangent space of the manifold X at the point x_o.

Let $v(w)$ denote the tangent vector containing the parametric curve w. For each real number k define the product of the real number k with the vector $v(w)$ by the equation

(Def) $$kv(w) = v(\omega)$$

where $\omega(t) = w(kt)$. This definition does not depend on which representative w is selected from the equivalence class $v(w)$. For let w_1 and w_2 be two parametric curves in $v(w)$ and let $\omega_i(t) = w_i(kt)$, $i = 1,2$. Then

$$\chi_{\omega_1}(\alpha) = k\chi_{w_1}(\alpha) = k\chi_{w_2}(\alpha) = \chi_{\omega_2}(\alpha) \quad .$$

The definition of addition of tangent vectors is more difficult. To facilitate our discussion, we define "coordinates" of tangent vectors. Let (f,U) be an allowable system of local coordinates at x_o. If $\alpha_E = \alpha f^{-1}$ and $\underline{w} = fw$, where $\underline{w} = (w^1, \dots, w^n)$, and if we suppose without loss of generality that α_E is continuously differentiable on $f(U)$, then

$\alpha w(t) = \alpha f^{-1} fw = \alpha_E\ (\underline{w}(t))$ for $|t|$ sufficiently small and

$$\frac{d}{dt}\left\{\alpha w(t)\right\}_{t=0} = \sum_{i=1}^{n} \frac{\partial\alpha_E}{\partial x^i}(x_o)\frac{dw^i}{dt}(0) \quad .$$

Since α and thus α_E run through all continuously differentiable functions, we may identify v(w) with the n-tuple $\underline{\dot{w}}(0)$ and say that $\dot{w}^i$, $i = 1, \ldots, n$, are <u>coordinates for v(w) in the system (f,U)</u>. It is clear that the coordinates of the tangent vector v are independent of the particular representative w selected from v. Given the n-tuple of real numbers $\underline{v} = (v^1, \ldots, v^n)$ and given the allowable system of local coordinates at x_o, (f,U), there is a tangent vector v with coordinates v^i, $i = 1, \ldots, n$, in the system (f,U). For let the functions w^i, $i = 1, \ldots, n$, be defined by the equations

$$w^i(t) = x_o{}^i + tv_1{}^i,$$

$i = 1, \ldots, n$, where $|t|$ is taken so small that $\underline{w}(t)$ lies in the open set f(U). Let $w(t) = f^{-1}(\underline{w}(t))$, and let $v = v(w)$. Then v is the desired tangent vector.

The preceding discussion establishes a one-to-one correspondence between triples (v,f,U) and triples $(\underline{v},f,U)$ where (f,U) is an allowable system of local coordinates at x_o, v is a tangent vector, and $\underline{v}$ is an n-tuple. Further, as v runs through all tangent vectors, the corresponding $\underline{v}$ runs through all n-tuples. This correspondence is indicated by

$$(v,f,U) \longleftrightarrow (\underline{v},f,U).$$

Let v_1 and v_2 be two tangent vectors at the point x_o. Addition of these two tangent vectors is defined by the equation

(Def) $$v_1 + v_2 = v$$

where

$$(v_1, f, U) \longleftrightarrow (\underline{v}_1, f, U) \quad ,$$

$$(v_2, f, U) \longleftrightarrow (\underline{v}_2, f, U) \quad ,$$

and

$$(v, f, U) \longleftrightarrow (\underline{v}_1 + \underline{v}_2, f, U) \quad .$$

We need to show that the vector v is determined independently of the allowable system (f,U). For this purpose we state the relation between any tangent vector v and the corresponding n-tuple in various allowable systems of coordinates.

Let v be a tangent vector to the manifold X at the point x_o. Let (f,U) and (g,V) be allowable local coordinates at x_o and

$$(v, f, U) \longleftrightarrow (\underline{v}, f, U)$$
$$(v, g, V) \longleftrightarrow (\underline{\tilde{v}}, g, V)$$

as explained in this subsection. Then the n-tuples $\underline{v}$ and $\underline{\tilde{v}}$ are related by the equations

$$\tilde{v}^i = \sum_{k=1}^{n} \frac{\partial \tilde{x}^i}{\partial x^k} (\underline{x}_o)\, v^k, \; i = 1, \ldots, n, \tag{1}$$

where

$$\underline{\tilde{x}} = gf^{-1}(x), \; \underline{x}_o = f(x_o) \text{ and } \underline{\tilde{x}}_o = g(x_o).$$

For a proof of this relationship see [CB; pp. 17-32] .

We apply equations (1) to show that the definition of the sum of two tangent vectors is independent of the allowable system of local coordinates used.

If

$$(v_1,f,U) \longleftrightarrow (\underline{v}_1,f,U) \quad ,$$

$$(v_2,f,U) \longleftrightarrow (\underline{v}_2,f,U) \quad ,$$

$$(v_1,g,V) \longleftrightarrow (\tilde{\underline{v}}_1,g,V) \quad , \text{ and}$$

$$(v_2,g,V) \longleftrightarrow (\tilde{\underline{v}}_2,g,V) \quad ,$$

then

$$\tilde{v}_1^i + \tilde{v}_2^i = \sum_{k=1}^{n} \frac{\partial \tilde{x}^i}{\partial x^k} (\underline{x}_o) \; (v_1^k + v_2^k) \quad ,$$

where $\tilde{\underline{x}} = gf^{-1}(\underline{x})$. Consequently, the definition of the sum of the tangent vectors v_1 and v_2 at the point x_o in X does not depend on which allowable system of local coordinates is used.

The set of all tangent vectors at the point x_o is, therefore, a real vector space. The dimension of this vector space equals the dimension of the underlying manifold (the dimension in both cases is n). This vector space is called the <u>tangent space of the manifold X at the point x_o</u>. It is also called the space of <u>contravariant vectors</u>.

2.2.4. <u>Tangent Vectors of Parametric Curves</u>

Let w be a parametric curve which is continuously differentiable on the open interval (a,b) and whose range is a subset of the differentiable manifold X of class C^p. Denote by $\dot{w}(t)$ the tangent vector to the parametric curve w at the point w(t). The function $\dot{w}$ is called the "instantaneous" velocity (vector) along the curve w. The value $\dot{w}(t)$ is called the velocity vector or the tangent vector to the curve w at the time t. It should be noted that in general the tangent vector $\dot{w}(t)$ lies in different tangent spaces for each time t. All of these tangent spaces are isomorphic [Hm; p. 15] , so for psychological reasons we think of $\dot{w}(t)$ as being in the same tangent space of all t.

Let t_o be a number in the open interval (a,b); let (f,U) be an allowable system of local coordinates at $w(t_o)$; and let $\underline{w}(t) = fw(t)$ for t in an open interval about t_o. Then the coordinates of $\dot{w}(t)$ with respect to the chart (f,U) for each t in this open interval about t_o are given by the n-tuple $\dot{\underline{w}}(t)$. Since w is continuously differentiable, the coordinates of $\dot{w}$ are continuous. Conversely, if the coordinates of $\dot{w}$ are continuous in an allowable system of coordinates, then w is continuously differentiable.

We say that the parametric curve w (in the manifold X of class C^p) is of class $C^q(a,b)$, q = 2, ..., p, if the coordinates of $\dot{w}$ are continuously differentiable of class C^{q-1} on (a,b). For q = 1 the definition has been given (see Subsection 2.2.2).

2.3. Covariant Vectors; Tensors

In this section X will denote an n-dimensional differentiable manifold with differentiable structure of class C^p defined by the atlas $\left\{(f_i,U_i)\right\}_{i=1}^{\infty}$. Let T_{x_o} denote the space of tangent vectors at the point x_o in X.

2.3.1. Dual Space, Covariant Vectors and their Coordinates

Let v^* be a real-valued, linear function (functional) with domain T_{x_o}. Denote by $T^*_{x_o}$ the collection of all such real-valued, linear functions. For v^*_1 and v^*_2 in $T^*_{x_o}$ and real numbers r_1 and r_2 define $r_1v^*_1 + r_2v^*_2$ by $(r_1v^*_1 + r_2v^*_2)(v) = r_1v^*_1(v) + r_2v^*_2(v)$ for any v in T_{x_o}. Then $T^*_{x_o}$ with this definition of addition and scalar multiplication is a real vector space. The space $T^*_{x_o}$ is called the dual space to the space T_{x_o}. The space $T^*_{x_o}$ is also called the space of covariant vectors at x_o and the members of $T^*_{x_o}$ are called covariant vectors. The vector space $T^*_{x_o}$ has dimension n, as we shall see. We drop the subscript x_o. However, all of our discussion is centered around the tangent space at an arbitrary but fixed point x_o in the manifold X.

We introduce coordinates for covariant vectors. Let (f,U) be an allowable system of local coordinates at the point x_o in X. Let the symbol δ^i_j, the Kronecker delta, be defined by

$$\delta^i_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j,\ i,j = 1, \ldots, n. \end{cases}$$

Let $\delta(j)$ (for integers $j = 1, \ldots, n$) be the vector in T which is defined by

$$(\delta(j),f,U) \longleftrightarrow (\underline{\delta}(j),f,U)$$

where $\underline{\delta}(j)$ is the n-tuple which has jth entry one and all other entries zero. The vectors $\delta(1), \ldots, \delta(n)$ form a basis for T [Hm; p.10]. If v is a vector in T and if $(v,f,U) \longleftrightarrow (\underline{v},f,U)$, then $v = \sum_{j=1}^{n} v^j\delta(j)$.

Let v^* be a member of T^*. Then $v^*(v) = \sum_{j=1}^{n} v^j v^*(\delta(j))$.

The values of v^* on the basis elements $\delta(1), \delta(2), \ldots, \delta(n)$ determine the values of v^* on T. Define the n real numbers v^*_j by

$$v^*_j = v^*(\delta(j)),\ j = 1, \ldots, n.$$

The real numbers v^*_j, $j = 1, \ldots, n$, are called the <u>coordinates of the covariant vector v^* in the system (f,U)</u>. Denote the n-tuple $(v^*_1, \ldots, v^*_n)$ by $\underline{v}^*$. As for the case of contravariant vectors, there is a one-to-one correspondence between the triples (v^*,f,U) and $(\underline{v}^*,f,U)$, where v^* is a covariant vector (at x_o), (f,U) is an allowable system of local coordinates (at x_o), and $v^*_1, v^*_2, \ldots, v^*_n$ are the coordinates of v^* in the system (f,U). This correspondence will be indicated by the grouping of symbols

$$(v^*,f,U) \longleftrightarrow (\underline{v}^*,f,U).$$

It is an evident consequence of the definitions of this section that if a and b are real numbers and if $(v^*,f,U) \longleftrightarrow (\underline{v}^*,f,U)$ and $(w^*,f,U) \longleftrightarrow (\underline{w}^*,f,U)$ for covariant vectors v^* and w^*, then

$$(av^* + bw^*,f,U) \longleftrightarrow (a\underline{v}^* + b\underline{w}^*,f,U).$$

If $(v^*,f,U) \longleftrightarrow (\underline{v}^*,f,U)$ and if $(v^*,g,V) \longleftrightarrow (\tilde{\underline{v}}^*,g,V)$, then

$$\tilde{v}^*_j = \sum_{k=1}^{n} \frac{\partial x^k}{\partial \tilde{x}^j}(\tilde{\underline{x}}_o)v^*_k,\ j = 1, \ldots, n, \tag{1}$$

where $\underline{x} = fg^{-1}(\tilde{\underline{x}})$. For a proof of equation (1) see [CB; pp. 17-32].

2.3.2. <u>Tensors (of Second Order)</u>

Let λ be a real-valued function on the product vector space $T \times T$. If the relations

$$\lambda(a_1v_1 + a_2v_2, w) = a_1\lambda(v_1, w) + a_2\ \lambda(v_2, w)$$

and

$$\lambda(w, a_1v_1 + a_2v_2) = a_1\ \lambda(w, v_1) + a_2\ \lambda(w, v_2)$$

hold for all real numbers a_1 and a_2 and for all contravariant vectors v_1, v_2 and w, then λ is called a <u>bilinear function</u> (functional) on $T \times T$.

Similarly, real-valued, bilinear functions on the product vector spaces $T \times T^*$, $T^* \times T$, and $T^* \times T^*$ may be defined. The set of real-valued bilinear functions on any one of these product vector spaces with evident definitions of scalar multiplication and addition forms an n^2-dimensional real vector space [Hm; pp. 36-37] . A bilinear function on $T \times T^*$ or $T^* \times T$ is called a <u>mixed tensor of second order</u>. A bilinear function on $T^* \times T^*$ is called a contravariant tensor of second order. A bilinear function on $T \times T$ is called a <u>covariant tensor of second order</u>. Tensors of higher order may be defined in an analogous fashion.

2.3.3. <u>Coordinates for Covariant Tensors (of Second Order)</u>

We restrict attention to covariant tensors (of second order). Let (f,U) be an allowable system of local coordinates and let the contravariant vector $\delta(j)$ for $j = 1, \ldots, n$ be defined by

$$(\delta(j), f, U) \longleftrightarrow (\underline{\delta}(j), f, U)$$

as before in Subsection 2.3.1. Suppose for contravariant vectors v and w

$$(v, f, U) \longleftrightarrow (\underline{v}, f, U)$$

and

$$(w, f, U) \longleftrightarrow (\underline{w}, f, U);$$

then

$$v = \sum_{j=1}^{n} v^j \delta(j) \quad \text{and} \quad w = \sum_{j=1}^{n} w^j \delta(j) \quad .$$

If λ is a covariant tensor of second order, then

$$\lambda(v,w) = \sum_{j=1}^{n} \sum_{k=1}^{n} v^j w^k \lambda(\delta(j), \delta(k)).$$

Let the n^2 real numbers λ_{jk} be defined by $\lambda_{jk} = \lambda(\delta(j), \delta(k))$ for $j,k = 1,2,\ldots, n$. Then

$$\lambda(v,w) = \sum_{j=1}^{n} \sum_{k=1}^{n} \lambda_{jk} v^j w^k \quad . \qquad (2)$$

The real numbers λ_{jk}, $j,k = 1, \ldots, n$, determine the values of λ for any pair of contravariant vectors v and w. These numbers are called the <u>coordinates of the covariant tensor λ</u> (of second order) <u>in the system (f,U)</u>. Symbolically, the relation between the covariant tensor λ and its coordinates λ_{ij}, $i,j = 1, \ldots, n$, is denoted by

$$(\lambda,f,U) \longleftrightarrow (\underline{\underline{\lambda}},f,U)$$

where $\underline{\underline{\lambda}}$ is the n by n matrix whose entry in the ith column and jth row is λ_{ij}.

Let (g,V) be a second allowable system of local coordinates (at x_o). Suppose that

$$(\lambda,g,V) \longleftrightarrow (\underline{\underline{\tilde{\lambda}}},g,V),$$

$$(v,g,V) \longleftrightarrow (\underline{\tilde{v}},g,V),$$

$$(w,g,V) \longleftrightarrow (\underline{\tilde{w}},g,V),$$

and

$$(\hat{\delta}(j),g,V) \longleftrightarrow (\underline{\delta}(j),g,V), \ j = 1, \ldots, n,$$

for the covariant tensor λ and the contravariant vectors v,w and $\hat{\delta}(j)$. Denote the (i,j) entry in the matrix $\underline{\underline{\tilde{\lambda}}}$ by $\tilde{\lambda}_{ij}$ (that is, $\lambda(\hat{\delta}(i),\hat{\delta}(j))$). Then the entries in the coordinate matrices $\underline{\underline{\tilde{\lambda}}}$ and $\underline{\underline{\lambda}}$ are related by the equations

$$\tilde{\lambda}_{jk} = \sum_{q=1}^{n} \sum_{r=1}^{n} \frac{\partial x^q}{\partial \tilde{x}^j}(\underline{\tilde{x}}_o) \frac{\partial x^r}{\partial \tilde{x}^j}(\underline{\tilde{x}}_o)\lambda_{qr}, \quad i,\ j = 1,\ 2,\ \ldots,\ n, \qquad (3)$$

where $\underline{x} = fg^{-1}(\tilde{x})$ (see [CB; pp 17-32]).

We note the evident fact that if λ is a covariant tensor (of second order) and if v is a fixed contravariant vector, then the linear functions obtained by inserting v as the first or as the second argument in λ (denoted by $\lambda(v,.)$ or $\lambda(.,v)$, respectively) are covariant vectors. If

$$(\lambda, f, U) \longleftrightarrow (\underline{\underline{\lambda}}, f, U)$$

and

$$(v, f, U) \longleftrightarrow (\underline{v}, f, U),$$

then

$$(\lambda(v,.), f, U) \longleftrightarrow (\underline{w}, f, U),$$

where

$$w_j = \sum_{i=1}^{n} \lambda_{ij} v^i, \quad j = 1, \ldots, n \quad .$$

2.3.4. <u>Coordinates of Contravariant and Mixed Tensors (of Second Order)</u>

Since the discussion concerning mixed and contravariant tensors (of second order) is analogous to that of covariant tensor (of second order), our account is abbreviated.

Let (f,U) and (g,U) be allowable systems of local coordinates. Let $\delta^*(j)$ and $\hat{\delta}^*(j)$ be covariant vectors such that

$$(\delta^*(j), f, U) \longleftrightarrow (\underline{\delta}(j), f, U)$$

and

$$(\hat{\delta}^*(j),g,V) \longleftrightarrow (\underline{\delta}(j),g,V)$$

for $j = 1,2,\ldots, n$, where $\underline{\delta}(j)$ is the n-tuple with jth entry one and all other entries zero. As before let $\delta(j)$ and $\hat{\delta}(j)$ be contravariant vectors so that $(\delta(j),f,U) \longleftrightarrow (\underline{\delta}(j),f,U)$ and $(\hat{\delta}(j),g,V) \longleftrightarrow (\underline{\delta}(j),g,V)$.

Since any statement concerning a mixed tensor on $T \times T^*$ has an obvious analogue concerning a mixed tensor on $T^* \times T$, we confine our discussion to the former case. Let λ be such a mixed tensor (of second order). By methods similar to those of Subsection 2.3.3, we may associate coordinates with the tensor λ. These coordinates are defined by the equations $\lambda^i_j = \lambda(\delta(i), \delta^*(j))$, i, j = 1, ..., n, in the system (f,U), and by the equations $\tilde{\lambda}^i_j = \lambda(\hat{\delta}(i), \hat{\delta}^*(j))$, i,j = 1, ..., n, in the system (g,V). The relations between the mixed tensor λ and its coordinates in each of the systems (f,U) and (g,V) are denoted by

$$(\lambda,f,U) \longleftrightarrow (\underline{\underline{\lambda}},f,U)$$

and

$$(\lambda,g,V) \longleftrightarrow (\tilde{\underline{\underline{\lambda}}},g,V),$$

where $\underline{\underline{\lambda}}$ and $\tilde{\underline{\underline{\lambda}}}$ are n by n matrices with (i,j) entries λ^i_j and $\tilde{\lambda}^i_j$ respectively.

The coordinates of the mixed tensor λ in the systems (f,U) and (g,V) are related by the equations

$$\tilde{\lambda}^i_j = \sum_{q=1}^{n} \sum_{r=1}^{n} \frac{\partial \tilde{x}^i}{\partial x^q}(\underline{x}_o) \frac{\partial x^r}{\partial \tilde{x}^j}(\tilde{\underline{x}}_o)\, \lambda^q_r \;, \quad i,j = 1, 2, \ldots, n \;, \qquad (4)$$

where $\underline{x} = fg^{-1}(\tilde{\underline{x}})$ (see [CB; pp 17-32]).

Let λ denote a contravariant tensor of second order. Coordinates may be associated with λ by methods similar to those discussed in the preceding subsections. The relations between the contravariant tensor λ and its coordinates in the systems (f,U) and (g,V) are denoted by

$$(\lambda, f, U) \longleftrightarrow (\underline{\underline{\lambda}}, f, U)$$

and

$$(\lambda, g, V) \longleftrightarrow (\underline{\underline{\widetilde{\lambda}}}, g, V),$$

where $\underline{\underline{\lambda}}$ and $\underline{\underline{\widetilde{\lambda}}}$ are n by n matrices with (i,j) entries λ^{ij} and $\widetilde{\lambda}^{ij}$ respectively. These coordinates are related by the equations

$$\widetilde{\lambda}^{ij} = \sum_{q=1}^{n} \sum_{r=1}^{n} \frac{\partial \widetilde{x}^{i}}{\partial x^{q}} (\underline{x}_o) \frac{\partial \widetilde{x}^{j}}{\partial x^{r}} (\underline{x}_o) \lambda^{qr}, \; i, \; j = 1, \; \ldots, \; n, \qquad (5)$$

where $\underline{\widetilde{x}} = gf^{-1}(\underline{x})$.

Covariant, contravariant and mixed tensors of order greater than two and the coordinates of such tensors may be defined in a fashion similar to the preceding. Relations may be established between coordinates of these tensors which are analogues of equations (3), (4), and (5). For a discussion of tensors see [CB; pp. 30-32] . We shall have no use in this work for such tensors.

2.4. Riemann Manifolds

To facilitate our definition and discussion of Riemann manifolds, we introduce additional concepts from topology.

Let A be a subset of the topological space X. Let Λ be a set, and suppose that to each element t in Λ there corresponds a subset G_t of X. Suppose that $A \subseteq \bigcup_{t \in \Lambda} G_t$; then the collection $\{G_t \mid t \in \Lambda\}$ is a cover of A. If each set G_t is open, the cover is said to be an open cover. We say that the set A is compact if every open cover of A has a finite subcollection which is also a cover of A.

We assume that X is a Hausdorff space with a countable base; the following propositions are well-known:

1. every sequence of points in a compact set has a subsequence which converges to a point of the set;
2. conversely, if every sequence of points in the set A has a subsequence which converges to a point of A, then A is compact;
3. if A is compact and if f is a continuous function with domain A, then the range f(A) is compact;
4. the closed interval of real numbers [a,b] is a compact subset of the real line in the usual topology.

For proofs of Propositions One, Two and Three, see [Pj, Teil I; pp. 83-90] ; for a proof of Proposition Four see [A; p.53].

2.4.1. Continuity of Vector- and Tensor-Valued Functions

Let λ be a function whose domain is the differentiable manifold X. Suppose that $\lambda(x)$ is a contravariant vector at x for each point x of the manifold X. We introduce an allowable chart (f,U) in a neighborhood of the point x_o. Define the function $\underline{\lambda}$ with domain U and range a subset of Euclidean n-space by

$$(\lambda(x),f,U) \longleftrightarrow (\underline{\lambda}(x),f,U)$$

(see Subsection 2.2.3).

Suppose that the manifold X is of class C^1 (C^p for integer p greater than one). We say that λ is continuous (of class C^q for q = 1, ..., p-1) at x_o if the function λ is continuous (of class C^q) on U for some allowable chart (f,U) at x_o. If the function λ is continuous (of class C^q) at each point of the manifold X, then we say that λ is continuous on X (of class C^q on X — $C^q(X)$).

The regularity properties of functions whose values are covariant vectors or whose values are tensors (of second order) may be defined in fashions analogous to that just given for contravariant vector-valued functions. We omit the precise formulation of these concepts.

2.4.2. Fundamental Form; Definition of Riemann Manifold

In Chapter One we have argued that Whittaker's theorem may be considered as a problem of Riemann geometry. Chapter Two is intended to make this argument more precise. One step in the link between Riemann geometry and Whittaker's theorem is the concept of the Riemann manifold.

For some positive integer p let X be a differentiable manifold of class C^p. Suppose that for each point x there is defined a symmetric, positive, covariant tensor $\lambda(x)$. That is, for all tangent vectors v and w which differ from zero

$$\lambda(x)(v,w) = \lambda(x)(w,v)$$

and

$$\lambda(x)(v,v) > 0.$$

If the function λ is continuous, then the set X with this function is called a Riemann manifold; the function λ is called the fundamental form of the manifold. We say that the Riemann manifold X is continuous (of class C^q) if its fundamental form is continuous (of class C^q).

If the contravariant vectors v and w have coordinates $v^1, \ldots, v^n$

and $w^1, \ldots, w^n$, respectively, and if $\lambda(x)$ has coordinates $\lambda_{ij}(x)$, $i,j = 1, 2, \ldots, n$, then

$$\lambda(x)(v,w) = \sum_{i,j=1}^{n} \lambda_{ij}(x)v^i w^j \quad .$$

Since $\lambda(x)(v,w) = \lambda(x)(w,v)$ for arbitrary contravariant vectors v and w, the coordinates of λ satisfy the relations

$$\lambda_{ij}(x) = \lambda_{ji}(x)$$

for all x in X. At each point x in the Riemann manifold X, the symmetric quadratic form

$$\sum_{i,j=1}^{n} \lambda_{ij}(x)v^i v^j$$

is positive definite. From linear algebra we know that there is a matrix which is inverse to the matrix $\underline{\underline{\lambda}}(x) = (\lambda_{ij}(x))$. Denote this inverse matrix by $\hat{\underline{\underline{\lambda}}}(x) = (\lambda^{ij}(x))$. Then

$$\sum_{j=1}^{n} \lambda^{ij}(x)\lambda_{jk}(x) = \sum_{j=1}^{n} \lambda^{ji}(x)\lambda_{kj}(x) = \delta^i_k$$

for $i,k = 1, \ldots, n$ and all x in X. It is easy to show the existence of a contravariant tensor function of second order, denoted by $\hat{\lambda}$, such that

$$(\hat{\lambda}(x), f, U) \longleftrightarrow (\hat{\underline{\underline{\lambda}}}(x), f, U),$$

where $\hat{\underline{\underline{\lambda}}}(x)$ is the n by n matrix $(\lambda^{ij}(x))$. The contravariant tensor $\hat{\lambda}$ is continuous (of class C^q) if λ is continuous (of class C^q).

2.4.3. <u>Length of a Parametric Curve</u>

Let x be a parametric curve of class C^q on the closed interval $[a,b]$. Since $\lambda(x(t))(\dot{x}(t),\dot{x}(t))$ is a non-negative real number for each t in $[a,b]$, it is meaningful to define the length of x by the equation

(Def)
$$L(x) = \int_a^b \sqrt{\lambda(x(t))(\dot{x}(t),\dot{x}(t))}\ dt$$
$$= \int_a^b \sqrt{\sum_{i,j=1}^{n} \lambda_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t)}\ dt,$$

where $\int_a^b$ denotes the Riemann integral over the interval $[a,b]$ and where λ_{ij} and $\dot{x}^i$, $i,j = 1, \ldots, n$, are coordinates of the fundamental tensor $\lambda(x)$ and the tangent vector $\dot{x}$ respectively. The real number $L(x)$ is called the <u>length of the parametric curve</u> x. Define the function s by

(Def)
$$s(\tau) = \int_a^\tau \sqrt{\sum_{i,j=1}^{n} \lambda_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t)}\ dt$$

for each real number τ in $(a,b]$ and

(Def) $\quad s(a) = 0.$

Then s is a continuous real-valued function which is non-decreasing on the interval $[a,b]$. If the function s is strictly increasing on $[a,b]$, then the parametric curve x is said to be <u>nowhere constant</u>.

In our account a careful distinction has been made between a tensor or vector and its coordinates with respect to an allowable chart. This distinction has resulted, necessarily, in complicated writing. To simplify our notation, we no longer distinguish in our writing between these concepts when no confusion should arise.

2.4.4. Curves of Minimal Length; Geodesics

Let X be a Riemann manifold of class C^q, $q \geqq 2$ (the differentiable manifold X is of class at least C^{q+1}). Let x be a nowhere constant parametric curve of class C^1 on the closed interval $[a,b]$. Suppose that the curve x satisfies the boundary conditions

$$x(a) = \overset{0}{x}$$

and

$$x(b) = \overset{1}{x} \quad .$$

If there is no curve joining the points $\overset{0}{x}$ and $\overset{1}{x}$ which is of class C^1 on $[a,b]$ and which has length less than the curve x, then we say that x is a minimal curve. The existence of a parametric curve between two fixed points which minimizes length is a problem in the calculus of variations. A necessary condition for the existence of such a curve is that a certain system of ordinary differential equations be satisfied.

In order to state this result precisely, we introduce some notation. Let f denote a real-valued function with domain the open subset of Euclidean (2n + 1) space

$$\left\{ (t,\underline{x},\underline{w}) \,\middle|\, -\infty < t < \infty \;,\; (\underline{x},\underline{w}) \in U \right\}.$$

Suppose that the function f has continuous partial derivatives. We denote the partial derivative $\frac{\partial f}{\partial x^i}$ by $f_{i|}$ and the partial derivative $\frac{\partial f}{\partial w^i}$ by $f_{|i}$ for each integer $i = 1, \ldots, n$. A function $\underline{x}$ of the real parameter t will be called admissible if it has the following properties:

1. $\underline{x}$ is defined, continuous and has a continuous derivative on the interval $[a,b]$,

2. $\underline{x}(a) = \overset{o}{\underline{x}}$ and $\underline{x}(b) = \overset{1}{\underline{x}}$ for fixed n-tuples $\overset{o}{\underline{x}}$ and $\overset{1}{\underline{x}}$, and

3. $(\underline{x}(t), \dot{\underline{x}}(t))$ belongs to the open set U for each time t in the interval $[a,b]$.

Lemma (Euler, Lagrange). If, for the fixed admissible function $\underline{x}$, the inequality

$$\int_a^b f(t,\underline{x}(t),\dot{\underline{x}}(t))\,dt \leqq \int_a^b f(t,\underline{z}(t),\dot{\underline{z}}(t))\,dt$$

is satisfied for an arbitrary admissible function $\underline{z}$, then the function $\underline{x}$ satisfies the system of ordinary differential equations (Euler's equations)

$$f_{i|}(t,\underline{x},\dot{\underline{x}}) - \frac{d}{dt} f_{|i}(t,\underline{x},\dot{\underline{x}}) = 0,\ i = 1,\ \ldots,\ n.$$

Before applying the Euler-Lagrange lemma to the problem of minimizing length, we parameterize the minimizing function in terms of length of arc. Let $\tau(s)$ denote the inverse function to the strictly increasing length-of-arc function $s(\tau)$ defined by

$$s(\tau) = \int_a^\tau \sqrt{\sum_{i,j=1}^{n} \lambda_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t)}\ \ dt.$$

Let the parametric curve y be defined by $y(s) = x(\tau(s))$ for each real number s in the interval $[0,L(x)]$. Then y is of class C^1 on $[0,L(x)]$ and is a nowhere constant curve of minimal length which joins $\overset{o}{x}$ to $\overset{1}{x}$. Furthermore, the curve y satisfies the condition

$$\sum_{i,j=1}^{n} \lambda_{ij}(y(s))\dot{y}^i(s)\dot{y}^j(s) = 1,$$

where $\dot{y}^i$ denotes the derivative of y^i with respect to its parameter.

By means of the Euler-Lagrange lemma, we may conclude that the curve y is of class C^2 on $[0,L(x)]$ and that it yields a solution to the differential equations

$$\ddot{y}^i(s) + \sum_{j,k=1}^{n} \Gamma^i_{jk}(y(s))\dot{y}^j\dot{y}^k = 0, \qquad i = 1, \ldots, n, \tag{1}$$

where

(Def) $$\Gamma^i_{jk}(y) = \sum_{r=1}^{n} \lambda^{ir}(y)\Gamma_{jk,r}(y)$$

and where

(Def) $$\Gamma_{jk,r}(y) = \tfrac{1}{2}\left[\frac{\partial\lambda_{rj}}{\partial y^k} + \frac{\partial\lambda_{kr}}{\partial y^j} - \frac{\partial\lambda_{jk}}{\partial y^r}\right]$$

for i, j, k, r = 1, 2, ..., n. The symbols $\Gamma_{jk,r}$ and Γ^i_{jk} are called the Christoffel symbols of the first and second kind respectively. Despite the suggestive notation, the Christoffel symbols do not yield the coordinates of a tensor.

Since

$$L(y) = \int_0^{L(x)} \sqrt{\sum_{i,j=1}^{n} \lambda_{ij}(y(s))\dot{y}^i\dot{y}^j}\, ds = \int_a^b \sqrt{\sum_{i,j=1}^{n} \lambda_{ij}(x(t))\dot{x}^i\dot{x}^j}\, dt = L(x),$$

a necessary condition for a nowhere constant, admissible parametric curve x to have minimal length is that the coordinates of the corresponding parametric curve y(s) satisfy equation (1). Any curve y(s) with parameter s defined as above is said to be <u>parameterized with respect to length of arc</u>.

We have seen that if the minimal curve y joins two fixed points and if it has length of arc as a parameter, then the coordinates of y necessarily satisfy equations (1) above. This necessary condition stated above is not a sufficient condition (for example — segments of great circles on a unit sphere having length exceeding π). A parametric curve whose coordinates satisfy equations (1) is called a geodesic curve.

We arrived at equation (1) by requiring y(s) to be parameterized with respect to length of arc. However, a geodesic curve need not have length of arc as a parameter. For example if y(s) is a geodesic curve, then so is w(t) where w(t) = y(ct) and c is any real number.

In any case a non-constant geodesic curve y(t) has the property that

$$\sum_{i,j=1}^{n} \lambda_{ij}(y(t))\dot{y}^i(t)\dot{y}^j(t) = K^2$$

for some real constant $K \neq 0$. This fact may be established by differentiating the left hand side of this equation with respect to t and using equation (1) to show that the derivative is identically zero. If K equals one, then t in absolute value is length of arc.

The concept of geodesic curve arose by seeking a curve of least length between two points. Such a curve must be a geodesic. However, as noted in this subsection, a geodesic curve need not have minimal length between its endpoints. In order to establish sufficient conditions for a curve to have least length, we investigate geodesics.

2.4.5. Fundamental Existence and Uniqueness Theorem for Solutions of Ordinary Differential Equations

Since geodesic curves are defined by the ordinary differential equations (1) of Subsection 2.4.4 and since our mechanical problem gives rise to ordinary differential equations, we present basic results concerning such equations and their solutions. Equations (1) of Section 2.4.4 may be written as

$$\frac{dy^i}{dt} = v^i$$
$$\frac{dv^i}{dt} = -\sum_{k,j=1}^{n} \Gamma^i_{jk}(y)v^j v^k \quad , \tag{2'}$$

$i = 1, \ldots, n$. Let $\underline{w}$ denote the 2n-tuple defined by

$$w^i = \begin{cases} y^i, \ i = 1, \ldots, n, \\ v^{i-n}, \ i = n+1, \ldots, 2n. \end{cases}$$

Let $\underline{F}$ be the 2n-tuple valued function which is defined by

$$F^i(\underline{w}) = \begin{cases} v^i, \ i = 1, \ldots, n, \\ -\sum_{j,k=1}^{n} \Gamma^{i-n}_{jk}(\underline{y})v^j v^k, \ i = n+1, \ldots, 2n. \end{cases}$$

With these conventions, equations (2') take the simplified form of

$$\dot{\underline{w}} = \underline{F}(\underline{w}) \ . \tag{2''}$$

As is well known from the theory of ordinary differential equations, we have the following:

Lemma 1. Let $\underline{F}(\underline{w})$ be a continuous function with domain E^m (Euclidean m-space) and range a subset of E^m. Let $\overset{o}{\underline{w}}$ be a point of E^m and b be a positive real number. Denote by $S(\overset{o}{\underline{w}},b)$ the set

$$\left\{ \underline{w} \;\middle|\; |w^i - \overset{o}{w}{}^i| < b, \ i = 1,2, \ldots, m \right\} .$$

Let the positive number A be selected so that $|F^i(\underline{w})| \leqq A$, $i = 1,2, \ldots, m$ for all $\underline{w}$ in $S(\overset{o}{\underline{w}},b)$. Suppose that $\underline{F}$ satisfies a Lipschitz condition on

$S(\overset{o}{\underline{w}},b)$; i.e., there is a positive number M so that

$$|F^i(\overset{1}{\underline{w}}) - F^i(\overset{2}{\underline{w}})| \leqq M \sum_{j=1}^{m} |\overset{1}{w}{}^j - \overset{2}{w}{}^j| \quad , \tag{3}$$

$i = 1, \ldots, m$, for all points $\overset{1}{\underline{w}}$ and $\overset{2}{\underline{w}}$ in $S(\overset{o}{\underline{w}},b)$. Then there exists exactly one continuously differentiable function $\underline{\phi}(t)$, defined for those real t with $|t| < b/A$, so that

$$\dot{\underline{\phi}}(t) = \underline{F}(\underline{\phi}(t))$$

and so that

$$\underline{\phi}(0) = \overset{o}{\underline{w}} \ .$$

The function $\underline{\phi}$ is called the solution to the differential equation

$$\dot{\underline{w}} = \underline{F}(\underline{w}) \tag{4a}$$

which satisfies the initial condition

$$\underline{w}(0) = \overset{o}{\underline{w}} \ . \tag{4b}$$

With evident significance, we write this function as $\underline{\phi}(t,\overset{o}{\underline{w}})$.

Since, for each point $\overset{1}{\underline{w}}$ in $S(\overset{o}{\underline{w}},b/2)$, the inequality $|F^i(\underline{w})| \leqq A$ holds for all points $\underline{w}$ in $S(\overset{1}{\underline{w}},b/2)$ and since F satisfies the same Lipschitz condition on $S(\overset{1}{\underline{w}},b/2)$, the function $\underline{\phi}(t,\overset{1}{\underline{w}})$ is defined for all $|t| < \frac{1}{2}(b/A)$. This argument shows that the solution $\underline{\phi}(t,\underline{w})$ is defined for all time t in an open interval $(-T,T)$ and all $\underline{w}$ in some neighborhood of $\overset{o}{\underline{w}}$.

Let θ be a real number between zero and one. For $|t| < \theta b/A$, the solution function $\underline{\phi}(t,\overset{o}{\underline{w}})$ is uniformly continuous in $\overset{o}{\underline{w}}$. That is, for any positive number ϵ, there is a positive number δ so that the inequalities $|\overset{1}{w}{}^i - \overset{o}{w}{}^i| < \delta$, $i = 1, \ldots, m$, imply that

$$|\phi^i(t,\overset{1}{\underline{w}}) - \phi^i(t,\overset{o}{\underline{w}})| < \varepsilon \ , \ i = 1, \ldots, m.$$

Other regularity properties of the solution function $\underline{\phi}(t,\overset{o}{\underline{w}})$ are contained in the

Lemma 2. If $\underline{F}$ is of class C^k on E^m, where k is a positive integer, then the solution function $\underline{\phi}$ is of class C^k in t and $\overset{o}{\underline{w}}$.

The proofs of these lemmas are widely available in the literature. For an elementary and readable account, see [BR; pp. 148-176]; for a more advanced but readable account, see [K; pp. 120-126, 141, 155-161]; for a modern treatment, see [H; pp. 12-24].

Applying these lemmas to the defining equations (1) of geodesic curves, we have the

Lemma 3. Let X be a Riemann manifold of class C^q, $q \geqq 2$; i.e., the coordinates of the fundamental form are of class C^q in the differentiable manifold of class at least C^{q+1}. Let $\overset{o}{x}$ be any point of X and $\overset{o}{v}$ be any tangent vector at $\overset{o}{x}$. Then there is exactly one geodesic $\phi(t,\overset{o}{x},\overset{o}{v})$ such that $\phi(0,\overset{o}{x},\overset{o}{v}) = \overset{o}{x}$ and $\dot{\phi}(0,\overset{o}{x},\overset{o}{v}) = \overset{o}{v}$. The function ϕ is of class C^q in t, $\overset{o}{x}$, and $\overset{o}{v}$.

2.5. Geodesics

In our intuitive discussion of Whittaker's theorem in Section 1.3 we described consequences of the hypotheses of this theorem in terms of convexity of a ring-shaped region. To do so, we required the result that any two neighboring points could be joined by a geodesic. For completeness, we give the arguments for this result.

2.5.1. Main Lemma Concerning Geodesics

In the previous section sufficient conditions were given for the existence and uniqueness of a geodesic emanating from a given point of a Riemann manifold and having a given direction at that point. Throughout this section, we assume these sufficient conditions are satisfied; that is, we denote by X a Riemann manifold of class C^2.

This section is devoted to proving the next lemma. The abstractions of Chapter Three are based on this lemma.

Lemma. Each point c of X has a neighborhood U with the property that any two points of U can be joined by a smooth curve of length not exceeding that of any other smooth curve joining the same two points.

The discussion of the previous section assumes the existence of minimal curves and concludes that the coordinates of such curves yield solutions to certain ordinary differential equations. Solutions to these ordinary differential equations were shown to exist (locally) and were called geodesics. An example was given to show that geodesics need not minimize length. The lemma of this subsection shows that in fact minimizing curves exist (locally).

Our proof is based on arguments given in [KN; pp. 149-151 and 166-167] and in [St; pp. 307-311]; the latter reference is more readable. As before let $\emptyset(t,x,v)$ denote the geodesic passing through x in the direction v. In coordinates at x we write $\underline{\emptyset}(t,\underline{x},\underline{v})$. The proof comes as a sequence of lemmas, which are interesting if taken by themselves.

2.5.2. Riemann Normal Coordinates

We suppose that an allowable system of local coordinates (f,U) is given in X and that f(U) is all of E^n. With our general assumptions, it follows that the function $\underline{\phi}(t,\underline{x},\underline{v})$ (geodesic through the point $\underline{x}$ with derivative $\underline{v}$ at $t = 0$) is defined for all time t with $|t| < T(\underline{x},\underline{v})$. The real number $T(\underline{x},\underline{v})$ depends on the points $\underline{x}$ and $\underline{v}$ — we may assume that $T(\underline{x},\underline{v}) = b/A$ where the numbers b and A are defined in Lemma 1 of Subsection 2.4.5.

Let $\overset{o}{\underline{x}}$ be given and denote by T_o the number $T(\overset{o}{\underline{x}},\underline{0})$ defined above so that the function $\underline{\phi}(t,\overset{o}{\underline{x}},\underline{0})$ is defined for $|t| < T_o$ (actually this function is defined for all time t in this case). But for $|x^i - \overset{o}{x}{}^i| < b/2$ and for $|v^i| < b/2$, it follows that the function $\underline{\phi}(t,\underline{x},\underline{v})$ is defined whenever $|t| < T_o/2$.

Let β denote a real number that differs from zero. The function $\underline{\Psi}(t) = \underline{\phi}(\beta t,\underline{x},\underline{v})$ is defined for all time t with $|t| < T_o/2|\beta|$, if $\underline{x}$ and $\underline{v}$ satisfy the inequalities stated above. Further, the function $\underline{\Psi}(t)$ is a geodesic (is a solution to equation (1) of 2.4.4) and satisfies the initial conditions

$$\underline{\Psi}(0) = \underline{x},$$

and

$$\dot{\underline{\Psi}}(0) = \beta\underline{v}.$$

By the uniqueness of such a solution we have

$$\underline{\Psi}(t) = \underline{\phi}(\beta t,\underline{x},\underline{v}) = \underline{\phi}(t,\underline{x},\beta\underline{v}).$$

If the number β is selected so that $|\beta| < T_o/2$, then $\underline{\phi}(1,\underline{x},\underline{v}\beta)$ is defined for all n-tuples $\underline{x}$ and $\underline{v}$ with

$$|x^i - \overset{o}{x}{}^i| < b/2$$

and

$$|v^i| < \tfrac{1}{2}b|\beta| .$$

The function $\underline{\emptyset}$ is used to introduce a special system of coordinates called Riemann Normal Coordinates (see [CB; pp. 114-116], [Py; pp. 291-293], [Pj, Teil 2; pp.65-68], or [St; pp. 307-309]).

Lemma 1. For any fixed n-tuple $\underline{\overset{o}{x}}$, the equation $\underline{z} = \underline{\emptyset}(1,\underline{\overset{o}{x}},\underline{v})$ defines a topological mapping between a neighborhood of $\underline{v} = \underline{0}$ and a neighborhood of $\underline{x} = \underline{\overset{o}{x}}$.

Proof. It suffices to show that the Jacobian determinant

$$\det\left(\frac{\partial z^i}{\partial v^j}\right)$$

is non-vanishing at $\underline{v} = \underline{0}$. But for integers $i,j = 1,2,\ldots, n$,

$$\begin{aligned}\frac{\partial z^i}{\partial v^j} &= \lim_{t\to 0} \frac{\emptyset^i(1,\underline{\overset{o}{x}}, t\underline{\delta}(j)) - \overset{o}{x}{}^i}{t} \\ &= \lim_{t\to 0} \frac{\emptyset^i(t,\underline{\overset{o}{x}},\underline{\delta}(j)) - \overset{o}{x}{}^i}{t} \\ &= \dot{\emptyset}^i(0,\underline{\overset{o}{x}},\underline{\delta}(j)) = \delta^i_j\end{aligned}$$

where $\underline{\delta}(j)$ is the n-tuple with ith entry δ^i_j. Evidently the Jacobian determinant has value one at $\underline{v} = \underline{0}$.

This lemma implies the existence of an open set $V(\overset{o}{x})$ containing any fixed point $\overset{o}{x}$ of the manifold X, with the property that if $\overset{1}{x}$ is a point of $V(\overset{o}{x})$, then it can be joined to $\overset{o}{x}$ by exactly one geodesic which lies in $V(\overset{o}{x})$. We call such a neighborhood $V(\overset{o}{x})$ a Riemann coordinate neighborhood, and we call the point $\overset{o}{x}$ the center of the neighborhood.

The Riemann coordinate neighborhood $V(\overset{o}{x})$ depends in general on its center $\overset{o}{x}$. In fact, there is a "uniform" neighborhood $U(\overset{o}{x})$ of any point $\overset{o}{x}$ so that if $\overset{1}{x}$ is a point of $U(\overset{o}{x})$, then the point $\overset{1}{x}$ is the center of a Riemann coordinate neighborhood which contains the open set $U(\overset{o}{x})$.

In order to establish the existence of the "uniform" neighborhood $U(\overset{o}{x})$, we investigate the equation $\underline{z} = \underline{\emptyset}(1,\underline{x},\underline{v})$. Lemma 1 tells us that this equation may be solved for $\underline{v}$ as a function of $\underline{z}$ for points $\underline{z}$ in a neighborhood of the point $\underline{x}$. The coordinates of the n-tuple $\underline{v}$ which corresponds to the point $\underline{z}$ in the Riemann coordinate neighborhood $V(\underline{x})$ are called the <u>Riemann normal coordinates centered at $\underline{x}$ of the point $\underline{z}$</u>. The n-tuple $\underline{v}$ corresponding to the point $\underline{z}$ depends on the center $\underline{x}$. The next lemma describes this dependence.

<u>Lemma 1'</u>. Let the mapping $(\underline{v},\underline{x}) \longrightarrow (\underline{z},\underline{w})$ be defined by the equations $\underline{z} = \underline{\emptyset}(1,\underline{x},\underline{v})$ and $\underline{w} = \underline{x}$. Then this mapping is topological between a neighborhood of $(\underline{0},\overset{o}{\underline{x}})$ and a neighborhood of $(\overset{o}{\underline{x}},\overset{o}{\underline{x}})$.

Proof. It suffices to show that the mapping $(\underline{v},\underline{x}) \longrightarrow (\underline{z},\underline{w})$ has a non-vanishing Jacobian determinant at $(\underline{0},\overset{o}{\underline{x}})$. In block form the Jacobian determinant of this mapping is

$$\det \begin{bmatrix} \left(\dfrac{\partial z^i}{\partial v^j}\right) & \left(\dfrac{\partial z^i}{\partial x^j}\right) \\ \\ \left(\dfrac{\partial w^i}{\partial v^j}\right) & \left(\dfrac{\partial w^i}{\partial x^j}\right) \end{bmatrix}_{(\underline{0},\overset{o}{\underline{x}})} = \det \begin{bmatrix} (\delta^i_j) & \left(\dfrac{\partial z^i}{\partial x^j}\right) \\ \\ (0) & (\delta^i_j) \end{bmatrix} = 1.$$

From the inverse function theorem, the lemma follows.

Suppose that the equations $\underline{z} = \underline{\phi}(1,\underline{x},\underline{v})$ and $\underline{w} = \underline{x}$ are solved for the variables $\underline{v}$ and $\underline{x}$ in terms of the variables $\underline{z}$ and $\underline{w}$. Denote the solution functions by $\underline{v}(\underline{z},\underline{w})$ and $\underline{x}(\underline{z},\underline{w})$, and denote the neighborhood of definition of these functions by $(U(\underline{\mathring{x}}),U(\underline{\mathring{x}}))$, where $U(\underline{\mathring{x}})$ is an open set containing the point $\underline{\mathring{x}}$.

From the form of the equations just solved it follows that for $\underline{w} = \underline{\mathring{x}}$, the function $\underline{v}(\underline{z},\underline{\mathring{x}})$ yields the Riemann normal coordinates of the point $\underline{z}$ (see Lemma 1). That is, the Riemann coordinate neighborhood $V(\underline{\mathring{x}})$ contains the open set $U(\underline{\mathring{x}})$. Since $\underline{v}(\underline{z},\underline{w})$ is defined for all points $\underline{z}$ and $\underline{w}$ in the open set $U(\underline{\mathring{x}})$, it is a suitable "uniform" neighborhood.

We list a few consequences of the preceding discussion.

Corollary 1. $\underline{v}(\underline{z},\underline{x})$ is a continuous function of the center x.

Corollary 2. Any point $\underline{\mathring{x}}$ has a neighborhood $U(\underline{\mathring{x}})$ (sufficiently small) so that $\underline{v}(\underline{z},\underline{x})$ is defined for all points $\underline{x}$ and $\underline{z}$ in $U(\underline{\mathring{x}})$.

Corollary 3. Any point $\underline{\mathring{x}}$ has a neighborhood $V(\underline{\mathring{x}})$ so that, if $\underline{z}$ is in $V(\underline{\mathring{x}})$, there is a geodesic joining $\underline{\mathring{x}}$ to $\underline{z}$. In particular, the geodesic

$$\underline{\phi}(t,\mathring{x},\underline{v}(\underline{z},\mathring{x}))$$

has the property $\underline{\phi}(0,\mathring{x},\underline{v}(z,\mathring{x})) = \underline{\mathring{x}}$ and $\underline{\phi}(1,\underline{\mathring{x}},\underline{v}(\underline{z},\underline{\mathring{x}})) = \underline{z}$.

Of course, the neighborhood $U(\underline{\mathring{x}})$of 2 is a subset of the neighborhood $V(\underline{\mathring{x}})$ defined in 3.

Riemann normal coordinates are not allowable in the sense defined in Subsection 2.2.1. However, the tensor transformation formulas which are valid for changes in allowable coordinates (see Section 2.3) are also valid for changes to Riemann normal coordinates. Since we do not use any of the properties of allowable coordinates in connection with Riemann normal coordinates, we omit the proof of this statement.

2.5.3. Geodesics as Shortest Curves

Let $\underline{z}$ be a point in a Riemann coordinate neighborhood centered at $\underline{x}$. Denote, as before, the Riemann coordinates of $\underline{z}$ by $v^1(\underline{z},\underline{x}), \ldots, v^n(\underline{z},\underline{x})$. Consider the geodesic

$$\phi(t,\underline{x},\underline{v}(\underline{z},\underline{x})) = \phi(t)$$

which is defined for t in the interval $[0,1]$ and which has, as end points, the points $\underline{x}$ and $\underline{z}$ (see Corollary 3 of Lemma 1').

The length of $\underline{\phi}$ is given by

$$\begin{aligned} L(\underline{\phi}) &= \int_0^1 \left\{ \sum_{i,j=1}^{n} \lambda_{ij}(\underline{\phi}(t))\dot{\phi}^i(t)\dot{\phi}^j(t) \right\}^{\frac{1}{2}} dt \\ &= \int_0^1 \left\{ \sum_{i,j=1}^{n} \lambda_{ij}(\underline{\phi}(0))\dot{\phi}^i(0)\dot{\phi}^j(0) \right\}^{\frac{1}{2}} dt \\ &= \left\{ \sum_{i,j=1}^{n} \lambda_{ij}(\underline{x})v^i(\underline{z},\underline{x})v^j(\underline{z},\underline{x}) \right\}^{\frac{1}{2}} , \end{aligned}$$

where λ_{ij} are the coordinates of the fundamental form λ, and where we have used the fact that $\sum_{i,j=1}^{n} \lambda_{ij}(\underline{\phi}(t))\dot{\phi}^i(t)\dot{\phi}^j(t)$ is a constant along geodesics (see Section 2.4).

The quantity

$$\sum_{i,j=1}^{n} \lambda_{ij}(\underline{x})v^i v^j$$

is a positive definite quadratic form. Define, for each positive number ρ, the set

$$S(\underline{x},\rho) = \left\{ \underline{z} \,\middle|\, \sum_{i,j=1}^{n} \lambda_{ij}(\underline{x})v^i(\underline{z},\underline{x})v^j(\underline{z},\underline{x}) < \rho^2 \right\} .$$

For sufficiently small ρ, $S(x,\rho)$ is an open subset of the Riemann coordinate neighborhood centered at $\underline{x}$. Define $\rho(\underline{x})$ as the supremum of values of ρ for which $S(\underline{x},\rho)$ is a subset of a Riemann coordinate neighborhood centered at $\underline{x}$.

Lemma 2. Let $0 < \rho < \rho(\underline{x})$ and let z be a point in $S(\underline{x},\rho)$. Then the geodesic

$$\emptyset(t,\underline{x},\underline{v}(\underline{z},\underline{x}))$$

has the shortest length of any continuously differentiable parametric curve joining $\underline{x}$ to $\underline{z}$.

A particularly simple proof of this lemma may be based on arguments found in [St; pp. 310-311]. The same arguments may be modified to yield the conclusion that no piecewise smooth parametric curve which joins the points $\underline{x}$ and $\underline{z}$ can have length smaller than the geodesic $\emptyset(t,\underline{x},\underline{v}(\underline{z},\underline{x}))$.

2.5.4. Proof of Main Lemma

We combine the lemmas of Sections 2.5.2 and 2.5.3 to prove the main lemma concerning geodesics (see Section 2.5.1).

From Subsection 2.5.2, any point $\overset{o}{\underline{x}}$ has a neighborhood $U_{\overset{o}{\underline{x}}}$ so that

$\underline{v}(\underline{z},\underline{x})$ is defined for all points $\underline{z}$ and $\underline{x}$ in $U_{\overset{o}{\underline{x}}}$ (i.e., any two points of $U_{\overset{o}{\underline{x}}}$ may be joined by a geodesic). There is no loss in generality in supposing that $U_{\overset{o}{\underline{x}}} = S(\overset{o}{\underline{x}}, \rho_o)$ for some positive number ρ_o.

The expression

$$\sum_{i,j=1}^{n} \lambda_{ij}(\underline{x}) v^i(\underline{z},\underline{x}) v^j(\underline{z},\underline{x})$$

is continuous in $\underline{z}$ and $\underline{x}$ for all $\underline{z}$ and $\underline{x}$ in $S(\overset{o}{\underline{x}}, \rho_o)$. Hence, there exists a positive number ρ_1, $3\rho_1 < \rho_o$, such that, if $\underline{x}$ is a point in $S(\overset{o}{\underline{x}}, \rho_1)$, then $S(\underline{x},2\rho_1)$ is an open subset of $S(\overset{o}{\underline{x}}, \rho_o)$. It follows from the validity of Lemma 2 for piecewise smooth curves that

$$S(\overset{o}{\underline{x}}, \rho_1) \subseteq S(\underline{x},2\rho_1).$$

For if $\underline{z}$ is a point of $S(\overset{o}{\underline{x}},\rho_1)$, it can be joined to $\underline{x}$ by a piecewise geodesic curve passing through the point $\overset{o}{\underline{x}}$ which has length less than $2\rho_1$. Since the geodesic joining the points $\underline{x}$ and $\underline{z}$ has length less than this piecewise geodesic, the point $\underline{z}$ must be in the set $S(\underline{x},2\rho_1)$. The proof of the lemma is complete.

The lemma and discussion of this section indicate that (locally) on X a distance function may be defined (length of geodesics). The notion of such a local distance will be investigated in the next chapter.

2.6. Newtonian and Lagrangian Dynamical Systems

In this section we explain more fully the connection between Newtonian dynamical systems and Riemann geometry (the Jacobi form of the Maupertuis principle).

2.6.1. Submanifolds of a Differentiable Manifold

Let X be an n-dimensional differentiable manifold of class C^q, $q \geqq 1$, and let Y be an open connected subset of X. In a natural way the differentiable structure on X induces a differentiable structure on Y. If (f,U) is an allowable system of local coordinates in X and if y is a point in the open set $U \cap Y$, then there is a neighborhood of y within $U \cap Y$ which is mapped by f onto an open ball in Euclidean n-space centered at f(y). Denote this neighborhood by V_y. Then we say that the chart (f,V_y) is an allowable system of local coordinates in the set Y. These allowable charts yield a differentiable structure of class C^q on the open set Y. We call this differentiable structure the induced structure on Y, and the differentiable manifold Y is called a submanifold of the manifold X. It is a consequence of their definitions that the tangent spaces to the manifolds X and Y at their common points are identical.

If X is a Riemann manifold of class C^q and if Y is an open subset of X, then the fundamental form on X induces in a natural way a fundamental form on Y. For if λ is the fundamental form on X, if y is a point of Y, if (f,U) is an allowable chart at y in X, if (f, V_y) is an allowable chart at y in Y, and if in the notation of 2.2.3

$$(\lambda,f,U) \longleftrightarrow (\underline{\underline{\lambda}},f,U)$$

and

$$(\lambda',f,V_y) \longleftrightarrow (\underline{\underline{\lambda}},f,V_y) \quad ,$$

then λ' is a fundamental form on Y. A curve in Y has the same length whether it is computed as a curve in Y or as a curve in X. The form λ' we call the induced fundamental form on the manifold Y. It is evident that the manifold Y with the induced fundamental form λ' is a Riemann manifold of class C^q.

For simplicity we write λ in place of λ' since they are identical on Y.

In a similar fashion any tensor or vector on the differentiable manifold X may be induced on the submanifold Y.

2.6.2. Newtonian and Lagrangian Dynamical Systems

We know that Newton's equations of motion (the equations which characterize a Newtonian dynamical system) by a re-numbering of coordinates and masses can be put in the form

$$\ddot{x}^i = -\frac{1}{m_i}\frac{\partial V}{\partial x^i}, \quad i = 1, \dots, mn,$$

for m particles in Euclidean n-space (see Section 1.1). If allowable coordinates which are not rectilinear are introduced in Euclidean n-space, then Newton's equations are transformed into the equations

$$\ddot{y}^i + \sum_{j,k=1}^{n} \Gamma^i_{jk}(y)\dot{y}^j\dot{y}^k = -\sum_{j=1}^{mn} \lambda^{ij}(y)\frac{\partial V}{\partial y^j}, \quad i = 1, \dots, mn .$$

This form of Newton's equations has the advantage that allowable transformations of coordinates yield equations which are again of the same form.

This "invariant" form of Newton's equations of motion suggests a generalization of our concept of a dynamical system. We give the precise nature of this generalization in the next few paragraphs.

Let X be an n dimensional Riemann manifold of class C^q, $q \geqq 2$, and suppose that X has fundamental form λ. Let x denote a parametric curve of class $C^2(a,b)$ with range a subset of X. Denote by $\mu(t_o)$ the contravariant vector, which is defined by

$$(\mu(t_o), f, U) \longleftrightarrow (\underline{\mu}(t_o), f, U),$$

where

$$\mu^i = \frac{D\dot{x}^i}{dt} = \ddot{x}^i + \sum_{j,k=1}^{n} \Gamma^i_{jk}(x)\, \dot{x}^j \dot{x}^i, \quad i = 1, \ldots, n.$$

That a contravariant vector $\mu(t_o)$ is unambiguously defined by these relations is a consequence of arguments in [Kr; pp. 216-218].

Let V be a real-valued function of class $C^2(X)$. The parametric curve x of class $C^2(a,b)$ is said to be a (conservative) Lagrangian dynamical system with potential V if

$$\frac{D\dot{x}^i}{dt}(t) = -\sum_{j=1}^{n} \lambda^{ij}(x(t))\frac{\partial V}{\partial x^j}(x(t)), \quad i = 1,\ldots,n, \tag{1}$$

for all t in (a,b). The Lagrangian dynamical system is said to have n degrees of freedom where n is the dimension of the manifold X. We shall prove Whittaker's theorem for Lagrangian dynamical systems.

Proposition 1 (Conservation of Energy). If a Lagrangian dynamical system x(t) has potential V then

$$\tfrac{1}{2}\lambda(x(t))\,[\dot{x}(t), \dot{x}(t)] + V(x(t)) = h = \text{const.}$$

Proof. Differentiate with respect to time and use equation (1) of this section and the definition of the Christoffel symbols to show the result is identically zero in t.

The real parameter h is determined by the initial conditions of the dynamical system and is called the energy (total energy) of the system.

Let X be a Riemann manifold with fundamental form λ of class C^q, $q \geqq 2$. Let V be a real-valued function of class $C^p(X)$, $p \geqq 2$. Consider the set

$$Y(h) = \left\{ x \varepsilon X \mid h - V(x) > 0 \right\} .$$

Suppose that the set $Y(h_o)$ is non-empty. Then it is open in X and each of its components is a submanifold of the differentiable manifold X. By component of $Y(h_o)$ we mean any subset which is connected and maximal (i.e., is not contained in any other connected subset of $Y(h_o)$). The expression

$$\tfrac{1}{2}\left[h_o - V(x)\right] \lambda (x)$$

is, for each point x in $Y(h_o)$, a positive definite covariant tensor of second order and defines a fundamental form on the open set $Y(h_o)$. This form is not induced on $Y(h_o)$ by the form λ on X. A component of the open set $Y(h_o)$ becomes a Riemann manifold of class $C^{\min\{p,q\}}$, where, by assumption, $\min\{p,q\} \geqq 2$.

2.6.3. Lagrangian Dynamical Systems and Maupertuis' Principle

In Section 1.1 we indicated a relation between Newtonian dynamical systems and geodesics in a Riemann manifold. As indicated in the previous subsection, Newtonian dynamical systems are special cases of Lagrangian dynamical systems. In this section we establish the relation indicated in Section 1.1 between a (conservative) Lagrangian dynamical system of energy h_o and the Riemann manifold $Y(h_o)$ with fundamental form $\tfrac{1}{2}\left[h_o - V(x)\right]\lambda(x)$. This relation is the subject of the next two propositions. These propositions are generalizations of the two propositions of

Subsections 1.1.2.

Proposition 1. Let $\emptyset(s)$ be a geodesic in the manifold $Y(h_o)$ (with fundamental form $\frac{1}{2}[h_o - V(x)]\lambda(x)$). Suppose that the parameter s is length of arc, i.e.,

$$\tfrac{1}{2}[h_o - V(\emptyset(s))]\ \lambda(\emptyset(s))[\dot{\emptyset}(s),\dot{\emptyset}(s)] = 1, \tag{2}$$

and that the real-valued, increasing and twice continuously differentiable function s(t) is defined by the differential equation

$$\frac{ds}{dt} = h_o - V(\emptyset(s))$$

and the initial condition $s(0) = 0$. Let the function $\Psi(t)$ be defined by

(Def) $\Psi(t) = \emptyset(s(t))$.

Then Ψ is the (conservative) Lagrangian dynamical system in X with potential V and energy h_o which satisfies the initial conditions

$$\Psi(0) = \emptyset(0)$$

$$\frac{d\Psi}{dt}(0) = \frac{d\emptyset}{ds}(0)\,[h_o - V(\emptyset(0))].$$

Proposition 2. Conversely, let Ψ be a (conservative) Lagrangian dynamical system with potential V(x) and energy h_o. Suppose that $h_o - V(\Psi(t)) > 0$ for all time t in the domain of Ψ and that Ψ satisfies the initial conditions

$$\Psi(0) = \overset{o}{x}$$

$$\frac{d\Psi}{dt} = \overset{o}{v}\ [h_o - V(\overset{o}{x})]$$

where $\overset{o}{v}$ is a contravariant vector. Let the real-valued function $t(s)$ be defined by

$$\frac{dt}{ds} = \frac{1}{h_o - V(\Psi(t))}$$

and the initial condition $t(0) = 0$. Define the function $\emptyset(s)$ by

(Def) $\qquad \emptyset(s) = \Psi(t(s))$.

Then $\emptyset(s)$ is a geodesic in $Y(h_o)$ which satisfies the initial conditions

$$\emptyset(0) = \overset{o}{x}$$

$$\dot{\emptyset}(0) = \overset{o}{v}$$

and the equation (2) above.

These two propositions constitute a precise statement of the Jacobi form of the Maupertuis principle.

Proof of Proposition 1. The proof we present comes from the lecture notes of a course given by Dr. R. Kurth.

From equation (2) and the fact that

$$\frac{d\emptyset^i}{ds} = \frac{d\Psi^i}{dt}\ \frac{dt}{ds}$$

we have the equation

$$1 = \tfrac{1}{2}\left[h_o - V(\emptyset(s))\right] \sum_{i,j=1}^{n} \lambda_{ij}(\emptyset(s))\dot{\emptyset}^i(s)\dot{\emptyset}^j(s)$$

$$= \tfrac{1}{2}\left[h_o - V(\Psi(t))\right] \sum_{i,j=1}^{n} \lambda_{ij}(\Psi)\frac{d\Psi^i}{dt}\ \frac{d\Psi^j}{dt}\left(\frac{dt}{ds}\right)^2$$

$$= \frac{h_o - V(\Psi(t))}{2\left(\frac{ds}{dt}\right)^2} \sum_{i,j=1}^{n} \lambda_{ij}(\Psi)\ \frac{d\Psi^i}{dt}\ \frac{d\Psi^j}{dt}$$

$$= \frac{1/2}{h_o - V(\Psi(t))} \sum_{i,j=1}^{n} \lambda_{ij}(\Psi)\ \frac{d\Psi^i}{dt}\ \frac{d\Psi^j}{dt}\ .$$

Thus,

$$\tfrac{1}{2}\lambda(\Psi(t))\left[\frac{d\Psi}{dt}, \frac{d\Psi}{dt}\right] + V(\Psi(t)) = h_o \tag{3}$$

(energy is conserved). Since the function $\emptyset(s)$ is a geodesic, it satisfies the equations

$$\ddot{\emptyset}^i(s) + \sum_{j,k=1}^{n} \tilde{\Gamma}^{\,i}_{jk}(\emptyset(s))\dot{\emptyset}^j(s)\dot{\emptyset}^k(s) = 0, \tag{4}$$

$i = 1, \ldots, n$, where $\tilde{\Gamma}^i_{jk}$ denotes a Christoffel symbol associated with the fundamental form on $Y(h_o)$

$$\tilde{\lambda}(x) = \tfrac{1}{2}[\,h_o - V(x)]\lambda(x).$$

Multiplying equation (4) by $\tilde{\lambda}_{mi}$ and summing with respect to the index i, we have the equations

$$0 = \sum_{i=1}^{n} \tilde{\lambda}_{mi}(\emptyset)\ddot{\emptyset}^i + \sum_{j,k=1}^{n} \tilde{\Gamma}_{jk,m}(\emptyset)\dot{\emptyset}^j\dot{\emptyset}^k, \tag{5}$$

$m = 1, \ldots, n$. But

$$\ddot{\emptyset}^i = \frac{\dfrac{d^2\Psi^i}{dt^2}}{[h_o - V(\Psi)]^2} + \frac{\dfrac{d\Psi^i}{dt}}{[h_o - V(\Psi)]^3}\sum_{r=1}^{n}\frac{\partial V}{\partial x^r}\frac{d\Psi^r}{dt}$$

and

$$2\tilde{\Gamma}_{jk,m}(x) = [h_o - V(x)]\,\Gamma_{jk,m}(x)$$

$$- \tfrac{1}{2}\left[\lambda_{mj}\frac{\partial V}{\partial x^k} + \lambda_{km}\frac{\partial V}{\partial x^j} - \lambda_{jk}\frac{\partial V}{\partial x^m}\right].$$

Substituting into (5) we have

$$0 = \sum_{i=1}^{n} \lambda_{mi} \frac{d^2\Psi^i}{dt^2} \frac{1}{h_o - V(\Psi)}$$

$$+ \frac{1}{[h_o - V(\Psi)]^2} \sum_{i,r=1}^{n} \lambda_{mi} \frac{d\Psi^i}{dt} \frac{\partial V}{\partial x^r} \frac{d\Psi^r}{dt}$$

$$+ \frac{1}{h_o - V(\Psi)} \sum_{j,k=1}^{n} \Gamma_{jk,m}(\Psi) \frac{d\Psi^j}{dt} \frac{d\Psi^k}{dt}$$

$$- \frac{1/2}{[h_o - V(\Psi)]^2} \sum_{j,k=1}^{n} \left[\lambda_{mj} \frac{\partial V}{\partial x^k} + \lambda_{km} \frac{\partial V}{\partial x^j} - \lambda_{jk} \frac{\partial V}{\partial x^m}\right] \frac{d\Psi^j}{dt} \frac{d\Psi^k}{dt}$$

$$= \frac{1}{[h_o - V(\Psi)]} \left\{ \sum_{i=1}^{n} \lambda_{mi} \frac{d^2\Psi^i}{dt^2} \right.$$

$$\left. + \sum_{j,k=1}^{n} \Gamma_{jk,m} (\Psi) \frac{d\Psi^j}{dt} \frac{d\Psi^k}{dt} + \frac{\partial V}{\partial x^m} \right\} ,$$

where we have used equation (3) to replace

$$- \frac{1/2}{[h_o - V(\Psi)]^2} \sum_{j,k=1}^{n} - \lambda_{jk} \frac{\partial V}{\partial x^m} \frac{d\Psi^j}{dt} \frac{d\Psi^k}{dt}$$

by

$$\frac{1}{[h_o - V(\Psi)]} \frac{\partial V}{\partial x^m} .$$

Thus,
$$\sum_{i=1}^{n} \lambda_{mi} \frac{d^2\Psi^i}{dt^2} + \sum_{j,k=1}^{n} \Gamma_{jk,m}(\Psi) \frac{d\Psi^j}{dt} \frac{d\Psi^k}{dt} = - \frac{\partial V}{\partial x^m} ,$$

and Ψ is a Lagrangian dynamical system with potential V and energy h_o.

Proposition Two is proved by similar methods, and we omit the details.

CHAPTER III

GEOMETRY OF DISTANCE SPACES

In Chapter Two a relation was given between dynamical systems and geodesics in Riemann manifolds. The role geodesic curves play in interpreting the hypotheses of Whittaker's theorem motivated the approach to this theorem through Riemann manifolds.

Vaguely speaking, the essential results of the previous chapter are the connection between geodesics of Riemann manifolds and dynamical systems, the local existence of a distance function (length of shortest curves) on such manifolds, and the local convexity of such manifolds (existence of shortest curves). These ideas and the role they play are emphasized in Chapter Three by their introduction as axioms in an abstract space. The remarks of Section 1.3 are important for motivating the approach in this chapter.

In Chapter Four we return to Riemann manifolds for generalizing Whittaker's boundary hypotheses.

3.1 Local Distance

A set X is called a local distance space if there is a family of pairs $\left\{(U_\tau, d_\tau) \,\middle|\, \tau\in T\right\}$ which satisfies the following conditions (T is an indexing set):

1. $\bigcup_{\tau\in T} U_\tau = X$,
2. d_τ is a distance function on U_τ (see Section 2.1), and

3. $d_s(x,y) = d_T(x,y)$ if s and y are points in $U_s \cap U_T$.

A subset G of the local distance space X is called open in X if $G \cap U_\tau$ is open in the distance space U_τ for each τ in T. The collection of open sets is a topology on X which is called the <u>topology of the local distance $\{d_\tau | \tau \in T\}$</u>.

Cordition 2 above guarantees that there is at most one number assigned by all the functions d_τ to any two fixed points x and y of X. This number (if existing) is called the <u>local distance</u> between x and y and is denoted by xy.

Any Riemann manifold X of class C^2 is a local distance space. For if x is any point of the manifold, if U_x is a neighborhood of this point such that a minimal curve joins any two of its points (see main lemma of Section 2.5), and if $d_x(y_1,y_2)$ is defined to be the length of this minimal curve joining the points y_1 and y_2 of U_x, then the family of pairs

$$\left\{ (U_x, d_x) \mid x \in X \right\}$$

is a local distance whose topology is identical with the existing topology on X.

A second example is obtained if X is a distance space with distance d. Let $T = \{1\}$, $U_1 = X$ and $d_1 = d$. Then, with the family $\{(U_1,d_1)\} = \{(x,d)\}$, the set X is a local distance space.

This section is devoted to proving a partial converse to this statement. A simple example shows that the unqualified converse is false. For let $X = \{x,y,z\}$, $U_1 = \{x,z\}$, $U_2 = \{y,z\}$, $d_1(x,z) = 1$, $d_2(y,z) = \frac{1}{2}$; then X with family $\{(U,d_1),(U,d_2)\}$ is a local distance space. The open sets in X are the empty set, X, $\{z\}$, and $\{x,y\}$. Since X is not a Hausdorff space, it is not a distance space (see Section 2.1).

3.1.1. Parametric Curves and Line Segments

Recall that a continuous function with domain a closed interval $[a,b]$ and range a subset of the topological space X is called a parametric curve in X. Let X be a local distance space, and suppose that there is a parametric curve f which joins the points x and y of X. Denote the domain of f by $[a,b]$, and suppose that the equality

$$f(t_1)f(t_2) = \frac{|t_1 - t_2|}{b - a} \; xy$$

is satisfied for each pair of real numbers t_1 and t_2 in $[a,b]$ (i.e., the local distance between $f(t_1)$ and $f(t_2)$ must be defined). Then the curve f is said to be a line segment. The point x is said to be joined to y by the line segment f. Evidently, if x is joined to y by a line segment, then y is joined to x by a line segment, and we say simply that x and y are joined by a line segment.

If each point of the local distance space X has an open neighborhood such that any pair of its points can be joined by a line segment, then we say that X is locally convex. Notice that the line segments are not required to lie in the neighborhood. For a different approach to this material, see [R; pp. 99-214].

It is evident that a Riemann manifold of class C^2 is locally convex (see Section 2.5). The third example of this section (see the last paragraph in the preamble to Section 3.1 above) shows that a local distance space need not be locally convex.

3.1.2. Length of a Parametric Curve

We first define length of parametric curves in the distance space X. Denote the distance between the points x and y of X by xy. Let f be a parametric curve in X with domain the closed interval [a,b].

By a partition of the interval [a,b] we understand a finite set of real numbers

$$\{t_o, t_1, \ldots, t_n\}$$

where $a = t_o < t_1 < \ldots < t_n = b$.

The length of the curve f is defined to be the non-negative real number (if existing) given by

$$\mathbb{L}(f) = \sup_{\Delta}\left\{\sum_{i=1}^{n} f(t_i)f(t_{i-1}) \mid \{t_o, t_1, \ldots, t_n\} = \Delta\right\}$$

where the supremum is taken over all partitions of the interval [a,b]. If the set of numbers

$$\left\{\sum_{i=1}^{n} f(t_i)f(t_{i-1}) \mid \{t_o, t_1, \ldots, t_n\} = \Delta, \Delta \text{ is a partition of } [a,b]\right\}$$

is not bounded, then we write $\mathbb{L}(f) = \infty$; otherwise, we write $\mathbb{L}(f) < \infty$.

Let $a < c < b$; it is evident that

$$\mathbb{L}(f) = \mathbb{L}(f|[a,c]) + \mathbb{L}(f|[c,b]) \tag{1}$$

where (for example) $f|[a,c]$ is the restriction of the function f to the interval [a,c].

The concept of a local distance space is too general to allow a reasonable definition of length of arc without some restrictions. As an example of possible pathology, the sets U_T which occur in the definition of local distance need not be open (see the example $X = \{x,y,z\}$ given in the preamble of this section).

In order to reflect more nearly the properties of Riemann manifolds, we consider only a restricted class of local distance spaces. These restrictions are embodied in the next definition. A connected local distance space is said to be <u>manifold-like</u> if the distance spaces U_τ are open for each τ in T and if for each point x in a set U_τ, there is a positive number δ_τ so that:

1. the closure in X of the ball $S_\tau(x,\delta_\tau) = \{y \mid d_\tau(x,y) < \delta_\tau\}$ is a subset of U_τ, and
2. the closure in X of $S(x,\delta_\tau)$ is compact.

We define length of arc in a manifold-like local distance space X. Let f be a parametric curve in X with domain $[a,b]$. Notice that property one in the definition of local distance space (see the beginning of Section 3.1) implies that $\{U_\tau \mid \tau \in T\}$ is a cover for X (an open cover since X is manifold-like). The range of f is a compact set so there is a finite cover of the form

$$U(1), U(2), \ldots, U(k)$$

where $U(1), \ldots, U(k)$ are members of the set $\{U_\tau \mid \tau \in T\}$.

The point f(a) is in one or more of the sets $U(1), \ldots, U(k)$, say in $U(1), \ldots, U(r_a), r_a \leqq k$. Since f is continuous and since the sets $U(1), \ldots, U(r_a)$ are open, $f[a,c)$ is a subset of each of these r_a subsets for $|a-c|$ sufficiently small. Let t_1 be the maximum value of c such that $f[a,c)$ is a subset of one of the sets $U(1), \ldots U(r_a)$. Evidently,

$f[a,t_1]$ is not a subset of any of the sets $U(1),\ldots,U(r_a)$ unless (possibly) $t_1 = b$. If $t_1 = b$, we stop the process. Otherwise, suppose that $t_1 < t_2 < \ldots < t_q < b$ have been selected. The point $f(t_q)$ is in one or more of the sets $U(r)$, $r = 1, \ldots, k$. Let t_{q+1} be the maximum value of t so that $f[t_q,t)$ is a subset of at least one of the sets $U(1),\ldots,U(k)$. If $t_{q+1} = b$, the process stops. If $t_n < b$ for each integer n, then we inductively define an increasing sequence of real numbers $\{t_n\}_{n=1}^{\infty}$. We show, in fact, that no such sequence can exist.

Let $t_o = \lim t_n$. The point $f(t_o)$ is in at least one of the sets $U(1), \ldots U(k)$, say $U(r_o)$. Since the function f is continuous, there is a positive integer N so that if $n \geqq N$, then $f[t_n,t_o] \subseteq U(r_o)$. Therefore, by definition of t_{n+1} we have $t_{n+1} > t_o$. This statement is inconsistent with the fact that the increasing sequence $\{t_n\}_{n=1}^{\infty}$ has limit t_o. Thus the sequence terminates after a finite number of steps, and there exist times

$$a = t_o < t_1 < \ldots < t_{n-1} < t_n = b$$

so that the image of the half-open interval $[t_{i-1},t_i)$, is a subset of one of the distance spaces $U(1),\ldots,U(k)$ and so that this image is maximal. The word maximal is understood to mean that the image of the closed interval $[t_{i-1},t_i]$ under the function f, $f[t_{i-1},t_i]$, is a subset of none of the sets $U(1),\ldots,U(k)$. The restriction of the function f to the interval $[t_{i-1},c]$, $c < t_i$, is a parametric curve in one of the distance spaces $U(1),\ldots,U(k)$, and, as such, this curve has a length which we denote by

$$\mathbb{L}(f|[t_{i-1},c]).$$

Equation (1) implies that $\mathbb{L}(f|[t_{i-1},c])$ is an non-decreasing function of c; let

$$\text{(Def)} \qquad \mathbb{L}_{i-1} = \lim_{c \to t_i -} \mathbb{L}(f|[t_{i-1},c]), \; i=1, \ldots, n.$$

The quantity $\mathbb{L}_{i-1}$ is a non-negative real number or infinity. Denote by $\mathbb{L}(f)$ the real number or infinity

$$\text{(Def)} \qquad \mathbb{L}(f) = \sum_{i=1}^{n} \mathbb{L}_{i-1} \; ,$$

where $\mathbb{L}(f) = \infty$ is understood if $\mathbb{L}_{i-1} = \infty$ for any $i = 1, \ldots, n$. We call $\mathbb{L}(f)$ the <u>length of the parametric curve f in the local distance space X</u>.

This definition of length of the parametric curve f appears to depend on the choice of the covering sets U(1), ..., U(k). Were there a genuine dependence, our definition would have little use. By means of the next proposition, which relates the length of the curve f to "sums over partitions of the interval [a,b]," we show that length of parametric curves does not depend on the finite covering U(1), ..., U(k).

<u>Proposition</u>. Let f be a parametric curve in the manifold-like local distance space X. Let the domain of f be the closed interval [a,b]. Then there is a partition

$$\Delta_1 = \{t_o, t_1, \ldots, t_m\}$$

of [a,b] so that

1. for each integer $j = 0,1,\ldots,m-1$ and for all t' and t" in the interval $[t_j, t_{j+1}]$, the local distance f(t')f(t") is defined, and

2. $\mathbb{L}(f) = \sup_{\Delta}\left\{ \sum_{k=0}^{n-1} f(s_k)f(s_{k+1}) \middle| \{s_o,\ldots,s_n\} = \Delta\right\}$
where the supremum is taken over all partitions $\Delta \geq \Delta_1$.

Proof. Let the partition $\Delta_o = \{t_o', t_1', \ldots, t_n'\}$ be selected as in the definition of $\mathbb{L}(f)$; i.e., $f[t_{i-1}', t_i')$ is a maximal subset of one of the sets U(1), ..., U(k). Since the parametric curve f is continuous, since the sets U(1), ..., U(k) are open, and since Δ_o is a finite set, there is a positive number δ_o so that $f(t_i' - \delta_o, t_i' + \delta_o)$ is a subset of one of the sets U(1), ..., U(k) for each integer i (the set $f(t_n' - \delta_o, t_n' + \delta_o)$ is to be interpreted as $f(t_n - \delta_o, t_n])$. We suppose that the number δ_o is chosen to satisfy the inequalities $\delta_o < t_i' - t_{i-1}'$ for integers i = 1, ..., n. Let the partition Δ_1 be given by

$$\Delta_1 = \{t_o', t_1' - \delta_o, t_1', t_2' - \delta_o, t_2', \ldots, t_n' - \delta_o, t_n'\} .$$

Then the partition Δ_1 evidently yields conclusion one of this proposition.

To establish conclusion two, we suppose that the curve f has finite length. As in the definition of length, let

$$\mathbb{L}_{i-1} = \lim_{c \to t_i^-} \mathbb{L}(f| [t_{i-1}, c]), \ i=1,\ldots,n.$$

Let ε be a positive number; there is a positive number δ_1 which we may suppose to be less than δ_o so that, if $0 < t_i - c_i < \delta_1$ for each integer i = 1,...,n, then

$$\mathbb{L}(f) = \sum_{i=1}^{n} \mathbb{L}_{i-1} \geq \sum_{i=1}^{n} \mathbb{L}(f| [t_{i-1}, c_i]) > \mathbb{L}(f) - \frac{\varepsilon}{3} \qquad (2)$$

and

$$\sum_{i=1}^{n} f(c_i)f(t_i) < \frac{\varepsilon}{3} . \tag{3}$$

From the definition of $\mathbb{L}(f \mid [t_{i-1},c_i])$, there exist partitions $\Delta(i)$ of $[t_{i-1},c_i]$, $i = 1,\ldots,n$, so that $t_i - \delta_o$ is in $\Delta(i)$ and so that

$$\sum_{i=1}^{n} \mathbb{L}(f \mid [t_{i-1},c_i]) \geqq \sum_{i=1}^{n} \sum_{\Delta(i)} f(t')f(t'')$$

$$\geqq \sum_{i=1}^{n} \mathbb{L}(f \mid [t_{i-1},c_i]) - \frac{\varepsilon}{3} . \tag{4}$$

By the sum $\sum_{\Delta(i)} f(t')f(t'')$, we understand the sum of elements of the form $f(t')f(t'')$ where t' and t'' are successive elements of the partition $\Delta(i)$. Let $\Delta = \bigcup_{i=1}^{n} \Delta(i)$; Δ is a partition of $[a,b]$ and

$$\sum_{\Delta} f(t')f(t'') = \sum_{i=1}^{n} \sum_{\Delta(i)} f(t')f(t'') + \sum_{i=1}^{n} f(c_i)f(t_i) \quad . \tag{5}$$

Combining inequalities (2), (3), and (4) with equation (5), we see that

$$\mathbb{L}(f) \leqq \sum_{i=1}^{n} \sum_{\Delta(i)} f(t')f(t'') + 2\varepsilon/3 \leqq \sum_{\Delta} f(t')f(t'') + 2\varepsilon/3. \tag{6}$$

We show that

$$\mathrm{IL}(f) \geqq \sum_{\Delta} f(t')f(t''). \tag{7}$$

This inequality, which in a distance space follows from the triangle inequality, is not so evident in a manifold-like local distance space.

$$\mathbb{L}_{i-1} = \lim_{c \to t_i-} \mathrm{IL}(f \mid [t_{i-1},c])$$

$$\mathrm{IL}_{i-1} = \mathbb{L}(f \mid [t_{i-1},c_i]) + \lim_{c \to t_i-} \mathrm{IL}(f \mid [c_i,c])$$

$$\mathbb{L}_{i-1} \geq \mathbb{L}(f \mid [t_{i-1}, c_i]) + \lim_{c \to t_i-} f(c_i)f(c)$$

$$\mathbb{L}_{i-1} \geq \mathbb{L}(f \mid [t_{i-1}, c_i]) + f(c_i)f(t_i)$$

Thus, $\mathbb{L}_{i-1} \geq \sum_{\Delta(i)}^{n} f(t')f(t'') + f(c_i)f(t_i)$,

and (7) follows from the definition of IL(f).

We have assumed that $\Delta \supseteq \Delta_1$. The combination of the inequalities (6) and (7) proves conclusion two for the case that $\mathbb{L}(f)$ is finite. The modifications for the non-finite case are evident.

Since the selection of the partition Δ_1 depends on the finite cover U(1),...,U(k), the length of the parametric curve f still appears to depend on this cover. We argue now that length is independent of the cover.

Let U'(1),...,U'(k') be a second finite cover of the range of the parametric curve f (as before, each of the sets U'(1),..., U'(k') is selected from the family of distance spaces $\{U_\tau \mid \tau \in T\}$ which occur in the definition of local distance space). Denote by IL'(f) the length (defined as above) of the curve f which depends on the cover U'(1),...,U'(k'). Let Δ_1' be a partition of $[a,b]$ so that

$$\mathbb{L}'(f) = \sup_{\Delta} \left\{ \sum_{\Delta} f(t')f(t'') \mid \Delta \supseteq \Delta_1' \right\} .$$

The symbols $\sum_{\Delta} f(t')f(t'')$ denote a sum of all elements of the form $f(t')f(t'')$ where t' and t" are successive times in the partition Δ. Let $\Delta'' = \Delta_1 \cup \Delta_i'$; then $\Delta_1'' \supseteq \Delta_1$ and $\Delta_1'' \supseteq \Delta_1'$. It is evident that

$$\mathbb{L}(f) = \sup_{\Delta} \left\{ \sum_{\Delta} f(t')f(t'') \mid \Delta \supseteq \Delta_1'' \right\}$$

and

$$\mathbb{L}'(f) = \sup_{\Delta}\left\{ \sum_{\Delta} f(t')f(t'') \,\middle|\, \Delta \supseteq \Delta_1'' \right\} .$$

Thus, $\mathbb{L}'(f) = \mathbb{L}(f)$, and length is shown to depend only on the curve f and not on a particular finite cover of its range.

A line segment f which joins the points x and y in the manifold-like local distance space X has length equal to the local distance between x and y, i.e., $\mathbb{L}(f) \doteq xy$. This statement is a consequence of the proposition above and the definition of a line segment.

A parametric curve f in the manifold-like local distance space X is called <u>rectifiable</u> if it has finite length.

The usual properties of length of curves may be proved for parametric curves in a manifold-like local distance space. For example, let f be a parametric curve with domain the closed interval $[a,b]$ and range a subset of the manifold-like local distance space X. Let c be a real number in the interval $[a,b]$. Then

$$\mathbb{L}(f) = \mathbb{L}(f \,|\, [a,c]) + \mathbb{L}(f \,|\, [c,b]).$$

This equation follows from the proposition of this subsection.

Suppose that the curve f is rectifiable. Let $s(t) = \mathbb{L}(f \,|\, [a,t])$; then $s(t)$ is finite for each t in the interval $[a,b]$. The function s is a continuous, non-decreasing, real-valued function. If the function s is strictly increasing, we say that the rectifiable curve f is <u>nowhere constant</u>.

If X is a Riemann manifold of class C^q, $q \geqq 2$, with local distance given by length of shortest curves, then there are two different definitions of length of a parametric curve. If the two definitions of length do not yield the same values, then it is possible that the two definitions of the term "nowhere constant" (see above and Subsection 2.4.3) yield different results. The next subsection is devoted to determining relations between the two concepts of length in a Riemann manifold.

3.1.3. Length in Riemann Manifolds of Class C^q, $q \geqq 2$.

Let X denote a Riemann manifold of class C^q, $q \geqq 2$. Then length of geodesics joining neighboring points gives a local distance on X. The local distance space X is manifold-like. Using this local distance, we may define length of parametric curves as described in the previous subsection. Denote by $\mathbb{L}(f)$ this new length of the parametric curve f. Let L(f) denote the length defined before by

$$\text{(Def)} \qquad L(f) = \int_a^b \left\{ \sum_{i,j=1}^{n} \lambda_{ij}(f(t))\dot{f}^i(t)\dot{f}^j(t) \right\}^{\frac{1}{2}} dt$$

(see Subsection 2.4.3).

We wish to compare the definitions $\mathbb{L}(f)$ and L(f). Let the times $a = t_o < t_1 < \ldots < t_{m-1} < t_m = b$ be selected so that a shortest curve joins $f(t_{i-1})$ to $f(t_i)$ for $i = 1,\ldots,m$. Let the parametric curve g be defined for t in the interval $[a,b]$ so that $g \mid [t_{i-1},t_i]$ is the aforementioned shortest curve joining $f(t_{i-1})$ to $f(t_i)$, $i = 1,\ldots,m$. We may assume, without loss of generality, that $g \mid [t_{i-1},t_i]$ is a geodesic as defined in Section 2.4 (this amounts to a particular choice of parameterization). Then $\dot{g}$ is continuous on each interval $[t_{i-1},t_i]$, $i = 1, \ldots, m$, and

$$L(f) = \sum_{k=1}^{m} \int_{t_{k-1}}^{t_k} \left\{ \sum_{i,j=1}^{n} \lambda_{ij}(f(t))\dot{f}^i(t)\dot{f}^j(t) \right\}^{\frac{1}{2}} dt$$

$$\geqq \sum_{k=1}^{m} \int_{t_{k-1}}^{t_k} \left\{ \sum_{i,j=1}^{n} \lambda_{ij}(g(t))\dot{g}^i(t)\dot{g}^j(t) \right\}^{\frac{1}{2}} dt$$

$$\sum_{k=1}^{m} f(t_{i-1})f(t_k).$$

The last inequality, combined with the proposition of the previous subsection, yields the inequality $L(f) \geqq \mathbb{L}(f)$. If the parametric curve f is nowhere constant in the sense of the previous subsection and if $L(f)$ is meaningful, then f is nowhere constant in the sense of Section 2.4.

Let g be a parametric curve with domain $[a,b]$. Suppose g joins the points x to y (i.e., $g(a) = x$ and $g(b) = y$) and has minimal $\mathbb{L}$-length; that is, $\mathbb{L}(g) \leqq \mathbb{L}(f)$ for any parametric curve f joining x to y. Let

$$s(t) = \mathbb{L}(g|[a,t])$$

and

$$G(s) = g(t(s)),$$

where $\qquad t(s_o) = \{t|s(t) = s_o\}$. The function G is defined and continuous on $[0, \mathbb{L}(g)]$ with $\mathbb{L}(G) = \mathbb{L}(g)$, $G(0) = x$, and $G(\mathbb{L}(g)) = y$.

Since the parametric curve G has minimal length, G satisfies the equation

$$G(s_1)G(s_2) = |s_1 - s_2|$$

for sufficiently small $|s_1 - s_2|$. To see this equation, we note that in any case $G(s_1)G(s_2) \leqq |s_1 - s_2|$ for sufficiently small $|s_1 - s_2|$. If strict inequality holds for all small $|s_1 - s_2|$, the curve G could be replaced by one of smaller length.

From the definition of $\mathbb{L}$-length and from the local distance on X it follows that the parametric curve G is a geodesic of the Riemann manifold. Therefore, the function $\dot{G}$ is continuous on $[0,L(g)]$, and

$$L(G|[s_1,s_2]) = G(s_1)G(s_2)$$

for $|s_1 - s_2|$ sufficiently small. Thus,

$$\mathbb{L}(G) = L(G).$$

These arguments have established

<u>Proposition 1</u>. If g is a parametric curve with domain $[a,b]$ and with least $\mathbb{L}$-length of all curves which join x to y in X, then there is a parametric curve G with domain $[0, \mathbb{L}(g)]$ for which the following equations hold:

$$G(0) = x = g(a),$$

$$G(\mathbb{L}(g)) = y = g(b),$$

$$\mathbb{L}(G|[0,s]) = s = L(G|[0,s]),$$

$$\mathbb{L}(G|[0,s(t)]) = s(t) = \mathbb{L}(g|[a,t]),$$

and $$\mathbb{L}(g) = L(G).$$

Furthermore, G is a geodesic of the Riemann manifold and

$$\sum_{i,j=1}^{n} \lambda_{ij}(G(s))\dot{G}^i(s)\dot{G}^j(s) = 1$$

for all s in $[0,L(G)]$.

From Proposition 1 and the inequality $\mathbb{L}(f) \leqq L(f)$, we have the

<u>Corollary</u>. Among all parametric curves joining the points x and y, the curve G has the least L-length.

In view of this corollary, finding a parametric curve of least L-length becomes a problem of finding a curve of least $\mathbb{L}$-length. It is the latter problem that we shall solve.

Agreement has been established between L- and $\mathbb{L}$-length only for special parametric curves, i.e., geodesics. In fact, a more general statement can be made.

Proposition 2. If g is a piecewise smooth parametric curve defined on $[a,b]$, then $L(g) = \mathbb{L}(g)$.

A proof may be constructed as in [R; pp. 135-138] . Since this proposition plays no role in our development, we omit its proof.

From Proposition 1 we see that if a curve is nowhere constant with respect to $\mathbb{L}$-length (i.e., nowhere constant as defined in this section), then it is nowhere constant with respect to L-length (as defined in Subsection 2.4.3)). Of course we assume that the curve is (piecewise) smooth. From Proposition 2 we see that the two definitions actually agree on (piecewise) smooth curves.

Since Proposition 2 has not been established, we shall henceforth understand the term nowhere constant in the sense defined in this section. As we have just argued, this new meaning is an extension and not a contradiction of the old one.

3.2. Intrinsic Distance

In this section we assume that the topological space X is a manifold-like local distance space which is also locally convex (see Sections 2.1 and 3.1), and we show that a distance can be defined on X which agrees with the local distance "in the small."

Under the general assumptions of this section, any two points of X may be joined by a rectifiable parametric curve. For let x be any point of X, and let A(x) denote the set of points of X which can be joined to x by a rectifiable curve. To prove the assertion, we need only show that the set A(x) is the space X.

The set A(x) is non-empty — in fact it contains a neighborhood of the point x (by local convexity of X). The set A(x) is open. For let y be a point of $A(x)$. Then by local convexity there is a neighborhood U of the point y such that any point of U can be joined to y by a line segment. Clearly, any point of U can be joined to x by a rectifiable curve passing through y. Thus, the neighborhood U is a subset of $A(x)$. Since the point y is arbitrary, the set $A(x)$ is open.

A similar argument shows that the set $A(x)$ is closed. But

$$X = [X - A(x)] \cup A(x)$$

with the sets $A(x)$ and $X\text{-}A(x)$ both being open. Since the space X is connected (see definition of manifold-like local distance spaces) and since the set $A(x)$ is non-empty, the set $X\text{-}A(x)$ must be empty. In other words, the set $A(x)$ contains the space X and the point x may be joined to any point y of X by a rectifiable curve. Since the point x was arbitrarily selected, we have established the assertion.

In a similar fashion we may show that any two distinct points of the space X may be joined by a nowhere constant rectifiable curve — even a nowhere constant polygonal curve. By a polygonal curve we mean a parametric curve composed of a finite number of line segments.

For each pair of points x and y in the local distance space X, let $d(x,y)$ denote the real number $\inf \{\mathbb{L}(f)\}$ where the infimum is taken over all curves joining the points x and y. The function d is a distance which we shall call the intrinsic distance on X. Alternately, we could have defined d as the infimum over all nowhere constant curves joining x and y or as the infimum over all polygonal curves joining x and y.

The only property of a distance function which is not, perhaps, evident is the triangle inequality (see Subsection 2.1.2); that is, for any three points x, y, and z of X

$$d(x,z) \leqq d(x,y) + d(y,z).$$

Let ε be a positive number; let the curves f and g be selected so that their domains are $[0,1]$ and $[1,2]$, respectively, with $f(0) = x$, with $f(1) = y = g(1)$, with $g(2) = z$, with $\mathbb{L}(f) < d(x,y) + \varepsilon/2$, and with $\mathbb{L}(g) < d(y,z) + \varepsilon/2$. Let the function F be defined by

$$F(t) = \begin{cases} f(t), & 0 \leqq t \leqq 1 \\ g(t), & 1 < t \leqq 2 . \end{cases}$$

Then, $d(x,z) \leqq \mathbb{L}(F) = \mathbb{L}(f) + \mathbb{L}(g)$ or

$$d(x,z) \leqq d(x,y) + d(y,z) + \varepsilon .$$

Since the positive number ε can be selected arbitrarily, the triangle inequality follows.

The next two propositions yield relations between the local distance on X and the corresponding intrinsic distance. Under the general assumptions of the section we have the

Proposition 1. For any fixed $\overset{\circ}{x}$ of X, the function of y, $d(\overset{\circ}{x},y)$, is continuous in y with respect to local distance. Further, if the local distance $\overset{\circ}{x}y$ is sufficiently small, then $d(\overset{\circ}{x},y) \leqq \overset{\circ}{x}y$.

Proof. We show that $d(\overset{\circ}{x},y)$ is continuous at the point $\overset{\circ}{y}$ in X. Let $U(\overset{\circ}{y})$ be a neighborhood of $\overset{\circ}{y}$ such that line segments join any pair of its points; the length of a line segment joining $\overset{\circ}{y}$ to y is $\overset{\circ}{y}y$ (see the definition of line segments). By definition of $d(\overset{\circ}{y},y)$ we have the inequality

$$d(\overset{\circ}{y},y) \leqq \overset{\circ}{y}y.$$

Since the function d is a distance function, we know that

$$\left| d(\overset{\circ}{x},\overset{\circ}{y}) - d(\overset{\circ}{x},y)\right| < d(\overset{\circ}{y},y).$$

The last two inequalities imply the proposition.

<u>Proposition 2</u>. Any point $\overset{\circ}{x}$ has a neighborhood, denoted here by $U(\overset{\circ}{x})$, so that, for each point y in $U(\overset{\circ}{x})$, we have the equality

$$d(\overset{\circ}{x},y) = \overset{\circ}{x}y.$$

Proof. Since X is manifold-like, the point $\overset{\circ}{x}$ has a neighborhood $S(\overset{\circ}{x},\delta_o)$ on whose closure the local distance is defined (see Property 1 in the definition of manifold-like local distance space in Subsection 3.1.2). Let y denote a point in $S(\overset{\circ}{x},\delta_o)$. If a parametric curve joining $\overset{\circ}{x}$ to y remains within the set $S(\overset{\circ}{x},\delta_o)$, the repeated application of the triangle inequality shows that the length of this curve is not less than the local distance $\overset{\circ}{x}y$. Otherwise, the curve will have a point in common with the boundary of $S(\overset{\circ}{x},\delta_o)$. By <u>boundary</u> of a set W we mean the set

(Def) $$\mathrm{bdry}(W) = \mathrm{cl}(W) \cap \mathrm{cl}(X - W),$$

where (recall) cl(W) denotes the closure of the set W.

If the curve joining $\overset{\circ}{x}$ to y has such a point in common with the set $\mathrm{bdry}(S(\overset{\circ}{x},\delta_o))$, then it has length exceeding δ_o (again by the triangle inequality). In both cases the curve has length not less than the local distance $\overset{\circ}{x}y$. Thus, for points y in the set $S(\overset{\circ}{x},\delta_o)$, we have the inequality

$$d(\overset{\circ}{x},y) \geqq \overset{\circ}{x}y.$$

Combined with Proposition 1, this inequality yields the assertion of Proposition 2.

We have still not used the compactness of the sphere $S(\overset{\circ}{x},\delta_o)$. This property does not play a role until later.

3.3 Distance Topology

In the previous section we introduced a distance in a local distance space which yields the same length as the local distance. In this section we investigate the relation between properties of distance spaces and existence of curves of least length.

3.3.1. Length and Distance Spaces

Let X be a distance space with distance between its points x and y denoted by xy. Let f denote a parametric curve in X with domain the closed interval $[a,b]$. Suppose that f is rectifiable with length denoted by $\mathbb{L}(f)$. Define as before (see Subsection 3.1.2) the continuous and non-decreasing function s by

$$s(t) = \mathbb{L}(f|[a,t]).$$

Let the function g be defined on the interval $[0,1]$ by

$$g(\tau) = f(t)$$

where

$$\tau\,\mathbb{L}(f) = \mathbb{L}(f|[a,t]).$$

The function g is a parametric curve in X and satisfies the Lipschitz condition

$$g(\tau_1)g(\tau_2) \leqq \mathbb{L}(f)|\tau_1 - \tau_2|.$$

The parametric curve g is called the <u>normal parameterization</u> of the curve f. As shorthand we denote the normal parameterization of the curve f by $\hat{f}$.

To see that the function g satisfies a Lipschitz condition, we argue only the case that the function f is nowhere constant. In this case the function $\tau(t)$ defined above is a continuous and strictly increasing function with range the interval $[0,1]$. Further, the function $g(\tau)$ is given by

$$g(\tau) = f(t(\tau))$$

where $t(\tau)$ is the continuous inverse to $\tau(t)$. Evidently,

$$g(\tau_1)g(\tau_2) = f(t_1)f(t_2) \leqq | s(t_1) - s(t_2)|$$

where $t_i = t(\tau_i)$, $i = 1, 2$. But

$$s(t_i) = \mathbb{L}(f)\tau_i, \; i = 1, 2;$$

so

$$g(\tau_1)g(\tau_2) \leqq \mathbb{L}(f)|\tau_1 - \tau_2| .$$

If f is not "nowhere constant," the argument is similar. There must be changes to account for the fact that $t(\tau)$ is not a real-valued function. Since the function f is indeed constant where $\tau(t)$ has no real-valued inverse function, a similar argument works in this second case.

Let $\mathbb{F}$ denote a family of parametric curves each having length not exceeding the positive real number M. We signify by $\hat{\mathbb{F}}$ the corresponding family of normal parameterizations. If τ_1 and τ_2 are real numbers in the closed interval $[0,1]$ with $|\tau_1 - \tau_2| < M^{-1}\epsilon$, then $f(\tau_1)f(\tau_2) < \epsilon$ for any curve $\hat{f}$ in $\hat{\mathbb{F}}$. This statement follows from the Lipschitz condition satisfied by normal parameterizations and the upper bound M.

Any family of functions $\mathbb{F}$ with domain the distance space X and range a subset of the distance space Y (both distances will be denoted by juxtaposition) is called <u>equicontinuous</u> if for any positive number ϵ, there is a positive number δ (depending only on ϵ) such that $f(x)f(z) < \epsilon$ whenever $xz < \delta$ (ϵ is independent of the function f).

In the preceding section we introduced a distance, called the intrinsic distance, on a connected, locally convex, manifold-like local distance space. This distance is the infimum of lengths of curves joining the points of the space. We ask if this infimum is realized for some curve. Our approach gives a partial answer if the points are sufficiently close (i.e., line segments). Similar questions have been answered (under appropriate hypotheses) by use of the theorem of Arzelà and Ascoli.

In order to state this theorem, we introduce additional concepts from distance topology.

3.3.2. Complete and Finitely Compact Distance Spaces.

A sequence of points $\left\{x_n\right\}_{n \in \tilde{S}}$ in the distance space X is called a Cauchy sequence if for any positive number ϵ, there is an integer N so that whenever integers n and m exceed N, the inequality $x_n x_m < \epsilon$ holds. If X is a local distance space, the definition is meaningful with the understanding that the distance between the points x_n and x_m must be defined for all sufficiently large integers n and m. In a distance space or a manifold-like local distance space, any converging sequence is a Cauchy sequence. If, conversely, any Cauchy sequence converges, we say that the distance space or local distance space is complete.

If a subsequence of a Cauchy sequence converges to a point, then evidently the sequence converges to the same point. From this statement and the alternate characterization of a compact topological space (in terms of sequences) given in Section 2.4, it follows that a compact space is complete.

A set B in a distance space X is called bounded if

$$\operatorname{Sup}\left\{xy \,\middle|\, x \in B, y \in B\right\} < \infty .$$

In Euclidean n-space a bounded, closed set is compact. This statement is not true in general distance spaces. However, Euclidean n-space is complete and manifold-like. It turns out that in any complete, manifold-like intrinsic distance space, a bounded closed set is compact. For a more general statement and proof see [R; p. 172]. Any distance space will be called finitely compact if all bounded closed sets are compact.

For simplicity we shall assume, where needed, that our distance space is finitely compact. Any compact distance space is finitely compact. As in Euclidean n-space, a finitely compact distance space is complete.

3.3.3. Distance Spaces of Functions — Parametric Curves

A function whose range is a bounded set will be called bounded. Let $\mathbb{F}$ be a family of bounded functions each having domain the set T and range a subset of the distance space X. For functions f and g in $\mathbb{F}$, let

(Def) $$fg = \sup\{f(t)g(t) \mid t \in T\}.$$

The mapping $(f,g) \to fg$ is a non-negative real-valued function. Since each function f and g is bounded, the triangle inequality implies that $fg < \infty$. This relation yields a distance on $\mathbb{F}$ called the distance of uniform convergence. If T is a compact distance space, if X is complete, and if C denotes the set of all continuous functions with domain T and with range a subset of X, then the space C with the distance of uniform convergence is complete (see [R; pp. 65-66]). A proof of this fact can be constructed as in analysis.

We require the following well-known

Lemma (Arzelà and Ascoli). Let T be a distance space with a countable base; let X denote a finitely compact space; and let $\mathbb{F}$ denote a family of continuous functions with domain T and range a subset of X. Then the following two statements are equivalent.

1. The closure of $\mathbb{F}$ is a compact subset of C (with the distance of uniform convergence).
2. The family $\mathbb{F}$ is equicontinuous and the set $\mathbb{F}(T)$,

   (Def) $\mathbb{F}(T) = \{y \mid y = f(t), f \in \mathbb{F}, t \in T\}$,

   is bounded.

For a proof of this theorem see [R; pp. 78-81]. The statement of this theorem in [R] is more general than that given above.

In view of the discussion in Subsections One and Two, we have the following

Corollary. Let X be finitely compact, and let $\mathbb{F}$ denote a family of parametric curves each having length not exceeding the positive number M. Suppose for each function f in $\mathbb{F}$ there is a real number t_f in the domain of f so that the set $\{f(t_f) \mid f \in \mathbb{F}\}$ is bounded. As before, denote by $\hat{\mathbb{F}}$ the family of normal parameterizations of curves in $\mathbb{F}$. Then $\hat{\mathbb{F}}$ has compact closure in the space of continuous functions with the distance of uniform convergence.

Proof. If N is a bound for the set $\{f(t_f) \mid f \in \mathbb{F}\}$, then M + N is a bound for $\{f(t) \mid f \in \mathbb{F}, t \in \text{domain } (f)\}$.

Thus M + N is a bound for $\hat{\mathbb{F}}[0,1]$. Recall that $\mathbb{F}$ is equicontinuous, and apply the Arzelà and Ascoli lemma.

3.3.4. Length of Parametric Curves as a Lower Semi-Continuous Function; Existence of Minimal Curves

With the distance of uniform convergence, let $C[a,b]$ denote the set of parametric curves in X which have domain the closed interval $[a,b]$. Let $\{f_n\}_{n=0}^{\infty}$ be a sequence of parametric curves in $C[a,b]$ with $\lim f_n = f_o$. Then $\underline{\lim}\ \mathbb{L}(f_n) \geqq \mathbb{L}(f_o)$ where $\underline{\lim}$ denotes limit inferior (the inequality is to be interpreted in an evident fashion if any of the quantities are not finite).

For let ϵ be a positive number, and let the partition of the closed interval $[a,b]$, $\Delta_\epsilon = \{t_o, \ldots, t_m\}$, be selected so that

$$\sum_{i=1}^{m} f_o(t_{i-1})f_o(t_i) > \mathbb{F}(f_o) - \epsilon/3$$

($> \frac{1}{\epsilon}$ if $\mathbb{F}(f_o)$ is not finite). Since the sequence $\{f_n(t_i)\}_{n=1}^{\infty}$ converges to $f_o(t_i)$ for each $i = 0, 1, \ldots m$, there is a positive integer N so that

$$\sum_{i=1}^{m} f_n(t_{i-1})f_n(t_i) > \sum_{i=1}^{m} f_o(t_{i-1})f_o(t_i) - \epsilon/3$$

for $n \geqq N$. Thus,

$$\mathbb{L}(f_n) \geqq \sum_{i=1}^{m} f_n(t_{i-1})f_n(t_i) > \mathbb{L}(f_o) - 2\epsilon/3$$

($>1/\epsilon - \epsilon/3$ if $\mathbb{L}(f_o)$ is not finite). Passing to the limit, we have

$$\underline{\lim}\ \mathbb{L}(f_n) > \mathbb{L}(f_o) - \epsilon$$

($>1/\epsilon - \epsilon/2$ if $\mathbb{L}(f_o)$ is not finite). Since ϵ is an arbitrary positive number, the result is a consequence of the last inequality.

A function f with domain a distance space T and range a subset of the extended real numbers is called <u>lower semi-continuous</u> on T if

$\underline{\lim}\ f(t_n) \geqq f(t_o)$ for any sequence $\{t_n\}_{n=0}^{\infty}$ with $\lim t_n = t_o$. We have argued that $\mathbb{L}$-length is a lower semi-continuous function.

If f is a lower semi-continuous function with domain the distance space T and if the set A is a compact subset of T, then there is a point t_o in the set A such that

$$f(t_o) = \inf\{f(t) \mid t \in A\}.$$

For let $L = \inf\{f(t) \mid t \in A\}$ and let the sequence of points from A, $\{t_n\}_{n=1}^{\infty}$, be selected so that

$$\lim f(t_n) = L.$$

Since A is compact, we may suppose without loss of generality that $\lim t_n = t_o$, and (therefore) t_o is a point of the set A. Thus,

$$L \leqq f(t_o) \leqq \underline{\lim}\ f(t_n) = \lim f\ (t_n) = L.$$

and the statement is proved.

In view of the corollary to the Arzelà and Ascoli lemma, if X is finitely compact and if the points x and y of X can be joined by a curve of finite length, say length M, then there is a curve joining x and y which has minimal length. For according to the corollary, the set of parametric curves

$$\mathrm{cl}\left\{\hat{f} \mid \hat{f}(0) = x,\ \hat{f}(1) = y,\ \mathbb{L}(\hat{f}) \leqq M\right\}$$

(recall that cl(A) denotes the closure of the set A), where $\hat{f}$ denotes normal parameterization, is compact. Since $\mathbb{L}$ is lower semi-continuous, and since the set of parametric curves given above is compact, there is a curve f_o in the set which has minimal $\mathbb{L}$-length. Such a curve we call a minimal curve.

3.3.5. Application to Manifold-like, Local Distance Spaces

In Section 3.2 we established that a manifold-like, local distance space which is locally convex may be viewed as a distance space with the intrinsic distance. That is, the intrinsic distance agrees with the local distance in the small. Additionally, if X is finitely compact in the intrinsic distance, we see from the previous subsection that there is a curve f joining any two points x and y of X so that

$$\mathbb{L}(f) = xy.$$

If $\hat{f}$ is the normal parameterization of the parametric curve f and if X is a Riemann manifold of class C^q, $q \geqq 2$, then $\mathbb{L}(f) = \mathbb{L}(\hat{f}) = L(\hat{f})$, where $L(\hat{f})$ is the length of $\hat{f}$ in terms of the fundamental form (see Subsection 3.1.3 for a discussion of the point). Recall the general relation, $\mathbb{L}(g) \leqq L(g)$. Thus, f is a geodesic in the manifold X. We have established the

Proposition. In a manifold-like, locally convex, local distance space which is finitely compact with respect to the intrinsic distance, any two points can be joined by a curve of minimal length and this length equals the intrinsic distance between the two points.

Corollary. Let X be a Riemann manifold of class C^q, $q \geqq 2$. Suppose that X is finitely compact with respect to the intrinsic distance. Then joining any two points of X there is a geodesic which has minimal length.

In essence we have assumed the existence of minimal curves in the small (our line segments) and demonstrated their existence in the large.

3.4 Unique Line Segments; Consequences

We assume that X is a connected, finitely compact, intrinsic distance space. We say that the point x is joined to the point y by a unique line segment if the normal parameterizations of any two line segments f and g, say $\hat{f}$ and $\hat{g}$, with $\hat{f}(0) = \hat{g}(0) = x$ and $\hat{f}(1) = \hat{g}(1) = y$ are identical. That is, $\hat{f}(t) = \hat{g}(t)$ for all t in the interval $[0,1]$. We denote the normal parameterization of the line segments joining x to y by $\emptyset(.,x,y)$. Since $\emptyset(.,x,y)$ is a normal parameterization of a line segment and since $IL(\emptyset) = xy$, it follows that

$$x\emptyset(t,x,y) = t(xy), \quad y\emptyset(t,x,y) = (1 - t)xy,$$

and

$$x\emptyset(t,x,y) + \emptyset(t,x,y)y = xy.$$

If to each point x of X there corresponds a positive number δ so that any two points of the ball $S(x,\delta)$ can be joined by a unique line segment, then we say that X has unique line segments in the small.

A Riemann manifold of class C^q, $q \geqq 2$, has unique line segments in the small. This statement has been strengthened in [HWi] for Riemann manifolds of dimension two.

When a space has unique line segments in the small, it follows as a consequence that the function $\emptyset(t,x,y)$ is a continuous function of the three variables t,x,y (if the distance xy is small). More precisely we have the

Proposition. Let X denote an intrinsic distance space which is finitely compact and which has unique line segments in the small (the normal parameterization of the unique line segment which joins the points x and y of X is denoted by $\emptyset(.,x,y)$ as above). Let $\{x_n\}_{n=0}^{\infty}$ and

$\{y_n\}_{n=0}^{\infty}$ be infinite sequences in X which converge to the points x_o and y_o respectively. Then

$$\lim \phi\ (.,x_n,y_n) = \phi(.,x_o,y_o)$$

in the uniform distance.

The proof is indirect. Suppose that the sequence $\{\phi(.,x_n,y_n)\}_{n=0}^{\infty}$ does not converge uniformly to $\phi(.,x_o,y_o)$. By definition of uniform convergence there exists a sequence of numbers $\{t_n\}_{n=0}^{\infty}$ in the interval $[0,1]$ and a subsequence $\{\phi(.,x_n,y_n)\}_{n\in S}$ so that

$$\phi(t_n,x_n,y_n)\phi(t_n,x_o,y_o) \geqq \delta$$

for all integers n in the set S and for some fixed positive number δ.

Since the closed interval $[0,1]$ is compact, we may suppose without loss of generality that the sequence $\{t_n\}_{n\in S}$ converges. From the relations

$$\mathbb{L}(\phi(.,x_n,y_n)) = x_n y_n \longrightarrow x_o y_o$$

and from the corollary to the theorem of Arzelà and Ascoli, it follows that there is a uniformly converging subsequence $\{\phi(.,x_n,y_n)\}_{n\in T}$ of the sequence $\{\phi(.,x_n,y_n)\}_{n\in S}$. Let f denote the limit curve; $f = \lim \phi(.,x_n,y_n)(n \in T)$. We show that f is the line segment $\phi(.,x_o,y_o)$.

The lower semi-continuity of curve length, the continuity of distance, and the fact that each curve $\phi(.,x_n,y_n)$ is minimal imply the relations

$$x_o y_o = \lim x_n y_n = \lim \mathbb{L}(\phi(.,x_n,y_n)) \geqq \mathbb{L}(f) \quad (n\in T).$$

But f is a curve joining x_o to y_o; so

$$\mathbb{L}(f) \geqq x_o y_o \quad .$$

These two inequalities imply that the curve f is minimal.

Applying the triangle inequality twice yields the inequality

$$x_o f(t) \leqq x_o x_n + x_n \emptyset(t,x_n,y_n) + \emptyset(t,x_n,y_n) f(t).$$

Since $\emptyset$ is a normal parameterization of a line segment,

$$x_n \emptyset(t,x_n,y_n) = (x_n y_n) t.$$

Thus,

$$x_o f(t) \leqq x_o x_n + t(x_n\ y_n) + \emptyset(t,x_n,y_n) f(t) \quad .$$

The first and last terms on the right hand side of the last inequality tend toward zero and the middle term tends toward $t(x_o y_o)$ as n runs through T. Thus,

$$x_o f(t) \leqq t(x_o y_o) \quad .$$

In a similar fashion we could show that

$$y_o f(t) \leqq (1 - t)(x_o y_o).$$

Combining these two inequalities, we see that

$$x_o f(t) + y_o f(t) \leqq x_o y_o \quad .$$

Therefore, $x_o f(t) + y_o f(t) = x_o y_o$ for all t in the interval [0,1]. Evidently, $x_o f(t) = t(x_o y_o)$ and $y_o f(t) = (1 - t)(x_o y_o)$.

Since the line segments are unique, the parametric curves f and $\emptyset(.,x_o,y_o)$ are identical. Thus,

$$\lim \emptyset(.,x_n,y_n)\emptyset(.,x_o,y_o) = 0 \quad (n \in T).$$

But the set T is a subset of the set S. We have argued that

$$\emptyset(t_n,x_n,y_n)\emptyset(t_n,x_o,y_o) \geqq \delta > 0$$

for all integers n in the set T (that is, in the set S). In view of the convergence of the sequence $\{t_n\}_{n\epsilon S}$ and in view of the uniform convergence of the sequence $\emptyset\{(\cdot,x_n,y_n)\}_{n\epsilon T}$, we have a contradiction.

3.5 Hadamard's Theorem

Whittaker's theorem, as given in Chapter One, was interpreted as an assertion of the existence of a non-trivial, closed geodesic curve. To describe these curves precisely, we introduce the concept of homotopy. If it is granted as true that Whittaker's hypotheses insure that the ring-shaped region is locally convex (see Section 1.3 and Chapter 4), then his theorem becomes a special case of a theorem of Hadamard.

In this section we prove a version of Hadamard's theorem which was discovered by this author before he became aware of a more general result in the literature (see [R]).

3.5.1. Homotopy

Let f and g be continuous functions each having domain the topological space Z and range a subset of the topological space Y. We say that f is homotopic to g if there is a continuous function $\Psi(z,t)$ of the two variables — z in the space Z and t in the interval $[0,1]$ — such that

$$\Psi(z,0) = f(z) \qquad \text{and} \qquad \Psi(z,1) = g(z) \quad .$$

If f is homotopic to g, then g is homotopic to f, and we say that f and g are homotopic. Homotopy is an equivalence relation on continuous functions with domain Z and range a subset of Y.

Let f and g be nowhere constant parametric curves in the distance space Y (recall that f and g must be rectifiable). Denote by $\hat{f}$ and $\hat{g}$, respectively, their normal parameterizations. Then necessary and sufficient conditions for f and g to be homotopic are that f and g have the same domain and that $\hat{f}$ and $\hat{g}$ be homotopic.

For suppose that the parametric curves f and g have the same domain, the interval $[a,b]$, and that the normal parameterizations $\hat{f}$ and $\hat{g}$ are homotopic. Denote by $\tau_f(t)$ and $\tau_g(t)$ the normal parameters, i.e., for example,

$$\tau_f(t) = \mathbb{L}(f|[a,t])/\mathbb{L}(f) \quad .$$

Since $\hat{f}$ and $\hat{g}$ are homotopic, there is a continuous function Ψ with domain the square $[0,1] \times [0,1]$ and range a subset of Y so that

$$\Psi(\tau,0) = \hat{f}(\tau)$$

and

$$\Psi(\tau,1) = \hat{g}(\tau)$$

for τ in the interval $[0,1]$. Let the continuous function Ψ' be defined on the rectangle $[a,b] \times [0,1]$ by

(Def) $$\Psi'(t,r) = \Psi(\tau_f(t),r).$$

Then $\Psi'(t,0) = \hat{f}(\tau_f(t)) = f(t)$ and $\Psi'(t,1) = \hat{g}(\tau_f(t))$.
Define the parametric curve h with domain $[a,b]$ by

(Def) $$h(t) = \hat{g}(\tau_f(t)).$$

Then the parametric curves f and h are homotopic. Let the continuous function Ψ'' be defined on the rectangle $[a,b] \times [0,1]$ by

(def) $$\Psi''(t,r) = \hat{g}\left[(1 - r)\tau_f(t) + r\tau_g(t)\right] .$$

Then $\Psi''(t,0) = h(t)$ and $\Psi''(t,1) = \hat{g}(\tau_g(t)) = g(t)$ so that h and g are homotopic. Since homotopy is an equivalence relation, our statement is verified in one direction.

The other direction can be verified by similar argruments — we omit the details.

The Whittaker theorem suggests the investigation of closed curves. For this investigation the concept of homotopy is not appropriate. That is, with this concept of homotopy a closed minimal curve of a homotopy class need not yield a periodic dynamical system. The minimal curve may have a "corner." The next definition (equivalent to that of [R]) will serve to eliminate "corners."

Let f and g be continuous functions from the closed interval $[a,b]$ with ranges in the distance space Y. Suppose that $f(a) = f(b)$ and $g(a) = g(b)$. We say that the function f is free-homotopic to the function g if there is a continuous function Ψ with domain the rectangle $[a,b] \times [0,1]$ and with range a subset of the distance space Y so that $\Psi(t,0) = f(t)$ and $\Psi(t,1) = g(t)$ for all t in $[a,b]$, and so that $\Psi(a,r) = \Psi(b,r)$ for all r in $[0,1]$. If f is free-homotopic to g, then g is free-homotopic to f, and we say that f and g are free-homotopic. Free-homotopy is an equivalence relation on closed parametric curves.

If the parametric curves f and g are nowhere constant, then necessary and sufficient conditions for these curves to be free-homotopic are that they have the same domain and that their normal parameterizations be free-homotopic. The arguments for this result are identical to those given for the similar assertion about homotopic curves.

In both the case of homotopy and the case of free-homotopy, the resulting equivalence classes of functions will be called homotopy classes. It should be clear from the context whether homotopy or free-homotopy is the relevant concept.

3.5.2. Homotopy and Local Convexity

The concept of (free-) homotopy is applied next to the problem of minimal closed curves in our locally convex, manifold-like, local distance space X. We assume that the space X is finitely compact with the induced intrinsic distance. We also assume that X has unique line segment in the small (see Section 3.4). The normal parameterization of the line segment joining the (sufficiently close) points x and y of the space X is denoted as before by $\emptyset(.,x,y)$.

Under the general assumptions of this section, we have the

Lemma. Let $\{f_n\}_{n=0}^{\infty}$ be a sequence of parametric curves each having as domain the interval $[a,b]$. Suppose that

$$\lim f_n = f_o$$

in the uniform distance. Then there exists a positive integer N so that the curve f_n is homotopic to the curve f_o for integers $n \geqq N$. If in addition $f_n(a) = f_n(b)$ for all integers n sufficiently large, then $f_o(a) = f_o(b)$ and there exists a positive integer N' so that the curves f_n and f_o are free-homotopic for $n \geqq N'$.

The next proposition is useful in proving this lemma.

<u>Proposition</u>. Let $\delta(x)$ be defined for each point x to be the supremum of the set of numbers δ such that any two points of the ball $S(x, \delta)$ can be joined by a unique line segment. Then $\delta(x)$ is a lower semi-continuous function.

Proof. It is evident that if the inequalities $xy<\delta(x)$ and $xz<\delta(x)$ are satisfied, then there exists a unique line segment which joins the points y and z. Consequently, if $\delta(x_o) = \infty$ for some point x_o in the space X, then $\delta(x) = \infty$ for all points x. In this case the lower semi-continuity is evident.

Suppose that $\delta(x)$ is finite for all points x. From the assumptions of this subsection, $\delta(x)$ is positive for each point of X. We argue the semi-continuity indirectly.

Suppose to the contrary that there is a sequence of points $\{x_n\}_{n=0}^{\infty}$ in X with $\lim x_n = x_o$ and $\underline{\lim}\ \delta(x_n) < \delta(x_o)$. We may suppose without loss of generality that $\lim \delta(x_n)$ exists. Let the positive number ϵ be selected so that $3\epsilon = \delta(x_o) - \lim \delta(x_n)$. There is a positive integer N so that $\delta(x_n) + 2\epsilon < \delta(x_o)$ for integers n exceeding N. Assume that the integer N is selected so large that $x_n x_o < \epsilon$. For $n \geqq N$, we have

$$S(x_n, \delta(x_n) + \epsilon) \subseteq S(x_o, \delta(x_o))$$

where $S(x,\rho) = \{y \mid xy < \rho\}$. Therefore, any two points of the ball $S(x_n, \delta(x_n) + \epsilon)$ can be joined by a unique line segment. This statement contradicts the definition of $\delta(x_n)$. Consequently, $\underline{\lim}\ \delta(x_n) < \delta(x_o)$ is impossible, and the proposition is proved.

Proof of the Lemma. We argue that the curve f_n is homotopic to the curve f_o for sufficiently large integers n. The case of free-homotopy is established in a similar fashion.

The range of the curve f_o is a compact set in X. From the lower semi-continuity of the function $\delta(x)$ on the space X, it follows that there is a real number t_o in the interval $[a,b]$ so that

$$\delta(f(t)) \geqq \delta(f(t_o))$$

for all real t in the interval $[a,b]$. Let $\delta_o = \delta(f(t_o))$; the number δ_o is positive.

From the uniform convergence of the sequence $\{f_n\}_{n=0}^{\infty}$, there is a positive integer N so that

$$f_n(t)f_o(t) < \delta_o$$

for all integers $n \geqq N$ and all real t in $[a,b]$. The definitions of δ_o and of the function $\delta(x)$ imply that the unique line segment $\emptyset(.,f_n(t), f_o(t))$ exists for each integer $n \geqq N$. Let the functions Ψ_n be defined by

(Def) $$\Psi_n(t,r) = \emptyset(r,f_n(t),f_o(t))$$

for $n \geqq N$. From the uniqueness of the line segments (see the proposition of Section 3.4), we know that the functions Ψ_n are continuous. Therefore, the curves f_n and f_o are homotopic for all integers n which exceed the integer N.

Corollary. The homotopy classes are closed.

3.5.2. Minimal Curves and Homotopy

Let X be a connected, finitely compact intrinsic distance space with unique line segments in the small.

Combining the arguments used in proving the existence of minimal curves (Section 3.2.4) with those of the previous subsection, we have the

<u>Theorem</u> (Hadamard). In each homotopy (free-homotopy) class there is a parametric curve whose length is less than or equal to the length of any other curve of this class.

If X is a Riemann manifold of class C^q, $q \geqq 2$, if X is finitely compact in the induced intrinsic distance, and if a minimal curve from a free-homotopy class has the normal parameterization $\emptyset$, then the curve $\emptyset$ is a restriction of a periodic geodesic to one period.

CHAPTER IV

ANALYTICAL CONDITION FOR LOCAL CONVEXITY AND WHITTAKER'S THEOREM

In this chapter the results of Chapter Three are applied to Riemann manifolds to prove Whittaker's theorem in a more general setting than that given in Chapter One. As in Chapter Two, λ will denote the fundamental form of the manifold. Similarly, other notations from Chapters Two and Three may be used without comment.

Throughout this chapter G will denote an open connected subset of a Riemann manifold X of class C^q, $q \geqq 3$. We suppose that the boundary of the open set G is given by

$$\text{bdry}(G) = \{x \mid f(x) = 0\} ,$$

where f is a real-valued function of class C^3 on an open set in the manifold X. We assume that the set bdry(G) is bounded and that the function f satisfies the inequality

$$\sum_{i,j=1}^{n} \lambda^{ij}(x)\frac{\partial f}{\partial x^i}(x)\frac{\partial f}{\partial x^j}(x) > 0$$

for all points x in bdry(G). Finally, the set $\text{cl}(G) = G \cup \text{bdry}(G)$ is assumed to be finitely compact in the intrinsic distance.

An analytical condition on f will be given which implies local convexity of the set $G \cup \text{bdry}(G)$. This condition is shown (in Section 4.4) to be a generalization of the Whittaker condition on the boundary of the ring-shaped region (see Chapter One). The local convexity and the theorem of Hadamard (see Section 3.5) combine to yield the Whittaker theorem. The first two sections of this chapter contain preliminary results.

4.1. Surface Normal Coordinates

This section is devoted to introducing a special system of coordinates. These coordinates are described in the next

Lemma. Let $\overset{o}{x}$ be a point of bdry(G). In a neighborhood of $\overset{o}{x}$ there is a system of local coordinates in which one of the coordinate parameters is in magnitude the length along a geodesic perpendicular to the boundary of G.

Proof. Since we have assumed that

$$\sum_{i,j=1}^{n} \lambda^{ij}(x)\frac{\partial f}{\partial x^i}(x)\frac{\partial f}{\partial x^j}(x) > 0,$$

at least one of the partial derivatives of f must be non-vanishing at the point $\overset{o}{x}$. Suppose with no loss of generality that at the point $\overset{o}{x}$, $\frac{\partial f}{\partial x^1}(\overset{o}{x})$ is different from zero. Denote by $\underline{w}/(1)$ the (n-1)-tuple $(w^2, w^3, \ldots, w^n)$ where $\underline{w}$ is the n-tuple $(w^1, w^2, \ldots, w^n)$.

Applying the implicit function theorem, we see that there is a real-valued function Ψ of $\underline{w}/(1)$ which is defined and continuous and has continuous partial derivatives in some neighborhood U_o^* of $\underline{\overset{o}{x}}/(1)$ in Euclidean (n-1) space E^{n-1} and satisfies the relations

$$\Psi(\underline{\overset{o}{x}}/(1)) = \overset{o}{x}{}^1 ,$$

and

$$f[\Psi(\underline{w}/(1)), \underline{w}/(1)] = 0$$

for all $\underline{w}/(1)$ in U_o^* .

Let $\underline{\nu}(\underline{w}/(1))$ denote the unit normal to the surface bdry(G) given by

(Def) $$\nu^i(\underline{w}/(1)) = \left.\frac{\sum_{j=1}^{n} \lambda^{ij}(\underline{x}) \frac{\partial f}{\partial x^j}(\underline{x})}{\sqrt{\sum_{p,q=1}^{n} \lambda^{pq}(\underline{x}) \frac{\partial f}{\partial x^q}(\underline{x}) \frac{\partial f}{\partial x^q}(\underline{x})}}\right|_{\underline{x} = [\Psi(\underline{w}/(1)), \underline{w}/(1)]} ,$$

$i = 1, \ldots, n$, and let $\emptyset(s,v,\nu)$ be the geodesic with $\emptyset(0,x,\nu) = x$ and $\dot{\emptyset}(0,x,\nu) = \nu$.

Let $\underline{\eta}(\underline{y})$ be the mapping defined by

(Def) $$\underline{\eta}(\underline{y}) = \underline{\emptyset}(y^1, [\Psi(\underline{y}/(1)), \underline{y}/(1)] , \underline{\nu}[\underline{y}/(1)]) .$$

At $y^1 = 0$ the Jacobian determinant is given by

$$\frac{\partial \underline{\eta}}{\partial \underline{y}} = \begin{vmatrix} \nu^1 & \underline{\nu}/(1) \\ -\tau(\underline{\nu}/(1))^* & \underline{\underline{I}} \end{vmatrix} ,$$

where

(Def) $$\tau = \tau(\underline{y}/(1)) = \frac{\sum_{i=1}^{n} \nu^i(\underline{y}/(1)) \frac{\partial f}{\partial x^i} [\Psi(\underline{y}/(1)), \underline{y}/(1)]}{\frac{\partial f}{\partial x^1} [\Psi(\underline{y}/(1)), \underline{y}/(1)]} ,$$

$$(\underline{\nu}/(1))^* = \begin{pmatrix} \nu_2 \\ \nu_3 \\ \cdot \\ \cdot \\ \cdot \\ \nu_n \end{pmatrix} , \qquad \nu_i = \sum_{j=1}^{n} \lambda_{ij} \nu^j, \quad i = 1, \ldots, n,$$

and $\underline{\underline{I}}$ is the (n-1) by (n-1) identity matrix.

To see that the Jacobian determinant is given by equation (1) above, we compute the partial derivatives

$$\left.\frac{\partial \eta^i}{\partial y^j}\right|_{y^1 = 0} ,$$ $i, j = 1, \ldots, n$. For integers $i = 1, \ldots, n$,

$$\left.\frac{\partial \eta^i}{\partial y^1}\right|_{y^1 = 0} = \dot{\emptyset}^i(0, [\Psi(\underline{y}/(1)), \underline{y}/(1)], \underline{\nu}[\underline{y}/(1)] = \nu^i(\underline{y}/(1))$$

by definition of the function $\underline{\emptyset}$. For integers j = 2, 3, ..., n,

$$\left.\frac{\partial \eta^1}{\partial y^j}\right|_{y^1=0} = \frac{\partial}{\partial y^j}\ \eta^1(0,\underline{y}/(1)) = \frac{\partial}{\partial y^j}\ \Psi\ (\underline{y}/(1))$$

$$= -\ \tau(\underline{y}/(1))\ \nu_j(\underline{y}/(1)).$$

For integers i,j = 2, ..., n,

$$\left.\frac{\partial \eta^i}{\partial y^j}\right|_{y^1=0} = \frac{\partial}{\partial y^j}\ \eta^i(0,\underline{y}/(1)) = \frac{\partial}{\partial y^j}\ (y^i) = \delta^i_j \quad ,$$

where we have used the fact that

$$\underline{\eta}(0,\underline{y}/(1)) = \underline{\emptyset}\left\{0, [\Psi\ (\underline{y}/(1)),\ \underline{y}/(1)]\ ,\ \underline{\nu}(\underline{y}/(1)\right\} = [\Psi(\underline{y}/(1)),\ \underline{y}/(1)]\ .$$

Computing the Jacobian determinant yields

$$\left.\frac{\partial \underline{\eta}}{\partial \underline{y}}\right|_{y^1=0} = \nu^1 + \tau\nu^2\nu_2 + \tau\nu^3\nu_3 + \ldots + \tau\nu^n\nu_n.$$

From the definition of τ ,

$$\left.\frac{\partial \underline{\eta}}{\partial \underline{y}}\right|_{y^1=0} = \frac{\nu^1 \frac{\partial f}{\partial x^1} + \left[\sum_{p=2}^{n} \nu^p\nu_p\right]\left[\sum_{i,j=1}^{n} \lambda^{ij}\ \frac{\partial f}{\partial x^i}\ \frac{\partial f}{\partial x^j}\right]}{\frac{\partial f}{\partial x^1}}$$

$$= \frac{\left[\sum_{p=1}^{n} \nu^p\nu_p\right]\left[\sum_{i,j=1}^{n} \lambda^{ij}\ \frac{\partial f}{\partial x^i}\ \frac{\partial f}{\partial x^j}\right]}{\frac{\partial f}{\partial x^1}}$$

$$= \sum_{i,j=1}^{n} \lambda^{ij}\ \frac{\partial f}{\partial x^i}\ \frac{\partial f}{\partial x^j} \Bigg/ \frac{\partial f}{\partial x^1} \neq 0.$$

Since the Jacobian determinant is not zero for any n-tuple $\underline{y}$ with $y^1 = 0$, the transformation $\underline{\mathbb{J}}$ is one-to-one and continuous in both directions from a neighborhood of the point $y^1 = 0$ ($y^2, \ldots, y^n$ arbitrary) to a neighborhood of the point on the surface

$$(\Psi(\underline{y}/(1)), \underline{y}/(1)).$$

The mapping $\underline{\mathbb{J}}^{-1}$ is one-to-one and continuous in a neighborhood of the point $\overset{o}{x}$ on the boundary of G. If x is in this neighborhood, the first entry in the n-tuple $\underline{\mathbb{J}}^{-1}(x)$ is in magnitude the length along a geodesic which is perpendicular to the surface bdry(G). The neighborhood may be taken so small that the first entry in $\underline{\mathbb{J}}^{-1}(x)$ is in magnitude the intrinsic distance between the point x and its "perpendicular projection" on the surface $[\Psi(\underline{x}/(1)), \underline{x}/(1)]$. In addition the neighborhood may be taken so small that the intrinsic distance between $\underline{\mathbb{J}}^{-1}(x)$ and its "perpendicular projection" $[\Psi(\underline{x}/(1)), \underline{x}/(1)]$ is the distance between x and the set bdry(G).

The coordinates given by $\underline{\mathbb{J}}^{-1}$ we call <u>surface normal coordinates</u>. The introduction of these coordinates is not allowable in general since they are not sufficiently differentiable. However, under our assumption on the function f(i.e., $f \epsilon C^3$) the coordinates will be "allowable" if the manifold X is regarded as a differentiable manifold of class C^2 (or a Riemann manifold of class C^1).

4.2. <u>An Analogue of the Theorem of the Mean of Differential Calculus</u>

In differential calculus the theorem of the mean is used to determine local behavior of real-valued functions which satisfy sufficient regularity conditions. We give a result which plays an analogous role for the real-valued function f. Recall the general assumptions of this chapter.

<u>Lemma</u>. Let g be a parametric curve with domain $[a,b]$. Suppose that g is in class C^q, $q \geqq 1$, and that the points g(a) and g(b) are on the boundary of the set G. Suppose also that the range of g is a subset of a neighborhood in which surface normal coordinates are introduced and in which the y^1-coordinate of any point in magnitude is the distance between the point and the boundary of G. Then, there is a time t_o in the open interval (a,b) such that

1. $g(t_o)\text{bdry}(G) \geqq g(t)\text{bdry}(G)$ for any t in $[a,b]$, and
2. there is a geodesic perpendicular to bdry(G) which intersects and is perpendicular to g at t_o and which has length equal to $g(t_o)\text{bdry}(G)$.

For a point x of X and a subset S we define the distance between x and S, xS, by

$$\text{(Def)} \qquad xS = \inf\{xy \mid y \in S\}.$$

The number xS is called the distance from the point x to the set S. By conclusion one of the lemma we assert the existence of a point on the parametric curve g having maximum distance from the set bdry(G).

Proof of the lemma. The real-valued function $d(t) = g(t)\text{bdry}(G)$ is continuous on the compact set $[a,b]$. Hence, there is a real number t_o between the end points a and b which (without loss of generality we may suppose) lies in the open interval (a,b) so that

$$g(t_o)\mathrm{bdry}(G) \geqq g(t)\mathrm{bdry}(G)$$

for any t in $[a,b]$. Let the n-tuple $\overset{o}{\underset{\sim}{y}} = (\overset{o}{y}{}^1, \ldots, \overset{o}{y}{}^n)$ yield surface normal coordinates for $g(t_o)$ where the magnitude of $\overset{o}{y}{}^1$ is the distance from $g(t_o)$ to the boundary of G.

The y^1 curve through the point $g(t_o)$ is normal to the boundary of G. We show that this curve is normal to the parametric curve g at $g(t_o)$.

Consider the displaced "parallel" surface $[\mathrm{bdry}(G)]^*$ defined by

$$[\mathrm{bdry}(G)]^* = \{\underset{\sim}{y} \mid y^1 = \overset{o}{y}{}^1\}$$

where $\underset{\sim}{y}$ is in surface normal coordinates. Since $g(t)\mathrm{bdry}(G) \leqq |\overset{o}{y}{}^1|$ for all t in $[a,b]$ (definition of $\overset{o}{\underset{\sim}{y}} = \underset{\sim}{g}(t_o)$) and since $g(t_o)\mathrm{bdry}(G) = |\overset{o}{y}{}^1|$, g is tangent to $[\mathrm{bdry}(G)]^*$ at $\overset{o}{\underset{\sim}{y}}$. Since y^1-curves are normal to $[\mathrm{bdry}(G)]^*$ at $\overset{o}{\underset{\sim}{y}}$, the y^1-curve through $g(t_o) = \overset{o}{\underset{\sim}{y}}$ is normal to g.

4.3. A Generalization of the Whittaker Condition

In his theorem (see Chapter One) Whittaker gives a condition on the boundary of a ring-shaped region which, he claims, implies the existence of periodic orbits of a dynamical system. In this section we give a condition on the boundary of our set G which implies local convexity of the set $G \setminus \mathrm{bdry}(G)$. In view of the connection between Riemann manifolds and dynamical systems (see Section 2.6) and in view of the connection between local convexity and the existence of minimal curves in free-homotopy classes, this condition implies the existence of periodic orbits (if the classes are non-empty). In the next section we shall show that this condition reduces to the Whittaker condition in the case considered by him.

Denote by $\overset{\circ}{\underline{\nu}} = (\overset{\circ}{\nu}{}^1, \ldots, \overset{\circ}{\nu}{}^n)$ and $\overset{\circ}{\underline{\nu}}{}^* = (\overset{\circ}{\nu}_1, \ldots, \overset{\circ}{\nu}_n)$, respectively, the contravariant and covariant vectors normal to the surface bdry(G) at the point $\overset{\circ}{x}$ (i.e., $\overset{\circ}{\nu}_i = c\frac{\partial f}{\partial x^i}(\overset{\circ}{x})$, i=1, ..., n, where

$$c = \left[\sum_{i,j=1}^{n} \lambda^{ij}(\overset{\circ}{x})\frac{\partial f}{\partial x^i}(\overset{\circ}{x})\frac{\partial f}{\partial x^j}(\overset{\circ}{x})\right]^{-\frac{1}{2}}$$

and

$$\overset{\circ}{\nu}{}^i = \sum_{j=1}^{n} \lambda^{ij}(\overset{\circ}{x})\overset{\circ}{\nu}_j \quad .$$

Suppose that for each sufficiently small positive t, the point $\phi(t,\overset{\circ}{x},\overset{\circ}{\nu})$ on the geodesic ϕ is a point in G (this requirement may make necessary the local replacement of the function f by the function -f; both the equation $-f(x) = 0$ and the equation $f(x) = 0$ define the set bdry(G)). Geometrically this condition signifies that the y^1-coordinates of points in G are positive while points not in the set $G\cup$bdry(G) have negative y^1-coordinates.

If the real-valued function $\alpha(t)$, defined for sufficiently small $|t|$ by $\alpha(t) = f(\phi(t,\overset{\circ}{x},\overset{\circ}{\nu}))$, is expanded by Taylor's theorem (i.e., $\alpha(t) = \alpha(0) + \dot{\alpha}(0)t + o(t)$), then

$$\begin{aligned}\alpha(t) &= \left[\sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(\overset{\circ}{x})\overset{\circ}{\nu}{}^i\right]t + o(t)\\ &= \left[\sum_{i,j=1}^{n} \lambda^{ij}(\overset{\circ}{x})\frac{\partial f}{\partial x^i}(\overset{\circ}{x})\frac{\partial f}{\partial x^j}(\overset{\circ}{x})\right]^{\frac{1}{2}} t + o(t),\end{aligned}$$

where o(t) is a function which after division by t tends to zero with t. We have used the facts that

$$\phi(0,\overset{\circ}{x},\overset{\circ}{\nu}) = \overset{\circ}{x}$$

and

$$\dot{\phi}(0,\overset{\circ}{x},\overset{\circ}{\nu}) = \overset{\circ}{\nu} \ .$$

From this expansion we conclude that the function f is positive for points in the set G and negative for points in the complement of the set $G \cup \mathrm{bdry}(G)$.

4.3.1. Convex Boundaries

Let $\underline{\mathring{v}}$ be any unit tangent vector to the surface bdry(G) at the point $\mathring{x}$; that is,

$$\sum_{i,j=1}^{n} \lambda_{ij}(\mathring{x})\mathring{v}^i\mathring{v}^j$$

vanishes. We suppose that surface normal coordinates have been introduced in a neighborhood of the point $\mathring{x}$.

Denote by $\theta(s,\mathring{x},\mathring{v})$ the "perpendicular projection" of the tangent geodesic $\phi(s,\mathring{x},\mathring{v})$ onto the surface bdry(G). That is, if the surface normal coordinates of the point $\underline{\phi}(s,\mathring{x},\mathring{v})$ are given by the n-tuple $[\phi^1,\underline{\phi}/(1)]$, then the n-tuple $[0,\underline{\phi}/(1)]$ yields the surface normal coordinates of the point $\theta(s,\mathring{x},\mathring{v})$. This "perpendicular projection" is defined for all sufficiently small numbers $|s|$. Note that

$$f(\theta(s,\mathring{x},\mathring{v})) = 0,$$

$\phi(0,\mathring{x},\mathring{v}) = \mathring{x} = \theta(0,\mathring{x},\mathring{v})$, and $\dot{\phi}(0,\mathring{x},\mathring{v}) = \mathring{v} = \dot{\theta}(0,\mathring{x},\mathring{v})$ (in surface normal coordinates $\dot{\phi}^1(0,\mathring{x},\mathring{v}) = \mathring{v}^1 = 0$).

We make the following

Definition. Suppose that for any unit tangent vector $\mathring{v}$ at $\mathring{x}$ the inequality

$$\sum_{i,j=1}^{n} \left[\ddot{\theta}^i(0,\mathring{x},\mathring{v}) - \ddot{\phi}^i(0,\mathring{x},\mathring{v})\right]\mathring{\nu}^j \lambda_{ij}(\mathring{x}) > 0 \qquad (1)$$

is satisfied; then we say that the boundary of G is convex at x toward its interior. Recall that the vector ν is the unit normal vector which points into G. If for each point $\mathring{x}$ on the set bdry(G) inequality (1) is satisfied, we say that the set <u>bdry(G) is a convex boundary</u>.

We relate inequality (1) above to an inequality involving the function f. Since $f\left[\theta(s,\overset{\circ}{x},\overset{\circ}{v})\right]$ vanishes identically in s,

$$0 = \frac{d}{ds}f\left[\theta(s,\overset{\circ}{x},\overset{\circ}{v})\right] = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\left[\theta(s,\overset{\circ}{x},\overset{\circ}{v})\right]\frac{d\theta^i}{ds}(s,\overset{\circ}{x},\overset{\circ}{v})$$

and

$$0 = \frac{d^2}{ds^2} f\left[\theta(s,\overset{\circ}{x},\overset{\circ}{v})\right] = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x^i \partial x^j}\left[\theta(s,\overset{\circ}{x},\overset{\circ}{v})\right]\dot{\theta}^i\dot{\theta}^j + \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\left[\theta(s,\overset{\circ}{x},\overset{\circ}{v})\right]\ddot{\theta}^i(s,\overset{\circ}{x},\overset{\circ}{v}).$$

Therefore,

$$\begin{aligned}
&\sum_{i,j=1}^{n} \left[\ddot{\theta}^i(0,\overset{\circ}{x},\overset{\circ}{v}) - \ddot{\phi}^i(0,\overset{\circ}{x},\overset{\circ}{v})\right]\overset{\circ}{v}{}^j\lambda_{ij}(\overset{\circ}{x}) \\
&= \sum_{i=1}^{n}\left[\ddot{\theta}^i(0,\overset{\circ}{x},\overset{\circ}{v}) - \ddot{\theta}^i(0,\overset{\circ}{x},\overset{\circ}{v})\right]\frac{\partial f}{\partial x^i}(\overset{\circ}{x}) \\
&= -\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x^i \partial x^j}(\overset{\circ}{x})\dot{\theta}^i(0,\overset{\circ}{x},\overset{\circ}{v})\dot{\theta}^j(0,\overset{\circ}{x},\overset{\circ}{v}) \\
&\quad + \sum_{i,j,k=1}^{n} \Gamma^i_{jk}(\overset{\circ}{x})\frac{\partial f}{\partial x^i}(\overset{\circ}{x})\dot{\phi}^j(0,\overset{\circ}{x},\overset{\circ}{v})\dot{\phi}^k(0,\overset{\circ}{x},\overset{\circ}{v}) \\
&= -\sum_{i,j=1}^{n} f_{|ij}(\overset{\circ}{x})\overset{\circ}{v}{}^i\overset{\circ}{v}{}^j
\end{aligned}$$

where for integers $i,j = 1, \ldots, n$,

(Def)
$$f_{|ij}(\overset{\circ}{x}) = \left[\frac{\partial^2 f}{\partial x^i \partial x^j}(\overset{\circ}{x}) - \sum_{k=1}^{n} \Gamma^k_{ij}(\overset{\circ}{x})\frac{\partial f}{\partial x^k}(\overset{\circ}{x})\right].$$

Hence the statement that bdry(G) is convex at $\overset{\circ}{x}$ toward the interior of G is equivalent to the assertion that

$$-\sum_{i,j=1}^{n} f_{|ij}(\overset{\circ}{x})\overset{\circ}{v}^i\overset{\circ}{v}^j$$

is a positive definite quadratic form for all tangent vectors $\overset{\circ}{v}$ to the surface at $\overset{\circ}{x}$. That is, the inequality (1) is equivalent to the inequality

$$-\sum_{i,j=1}^{n} f_{|ij}\,(\overset{\circ}{x})\overset{\circ}{v}^i\overset{\circ}{v}^j > 0 \qquad (1')$$

for all $\overset{\circ}{v}$ such that

$$0 = \sum_{i,j=1}^{n} \lambda_{ij}(\overset{\circ}{x})\overset{\circ}{v}^i\overset{\circ}{v}^j = \sum_{i=1}^{n} \overset{\circ}{v}^i \frac{\partial f}{\partial x^i}(\overset{\circ}{x}) \quad .$$

Notice also that our definition of convexity of bdry(G) at $\overset{\circ}{x}$ toward the interior of G is independent of the choice of allowable coordinates which we introduce at the point $\overset{\circ}{x}$. This statement can be seen directly from inequality (1) of the previous section; or, by using the transformation rules for the Christoffel symbols, we could show that there is a covariant tensor of second order which in any allowable chart has coordinates

$$f_{|ij}(\overset{\circ}{x}),\ i,j = 1,\ \ldots,\ n,$$

at the point $\overset{\circ}{x}$.

4.3.2. Local Convexity of the Set G

Under the general assumptions of this chapter we have the

Proposition. If G has a convex boundary at $\overset{\circ}{x}$, then for sufficiently small values of the parameter $|s|$ the geodesic tangent at $\overset{\circ}{x}$ to the set bdry(G), $\emptyset(s,\overset{\circ}{x},\overset{\circ}{v})$, has no points in common with the set G.

Proof. Let the real-valued function α be defined for sufficiently small values of $|s|$ by

(Def) $\alpha(s) = f(\emptyset(s,\overset{o}{x},\overset{o}{v})),$

where $\emptyset(s,\overset{o}{x},\overset{o}{v})$ is a geodesic tangent to the set bdry(G) at $\overset{o}{x}$. By Taylor's formula with remainder

$$\alpha(s) = \alpha(0) + \dot{\alpha}(0)s + \ddot{\alpha}(0)s^2/2 + o(s^2) \ ,$$

where $o(s^2)$ denotes a function which after division by s^2 tends to zero with s.

Evaluating $\alpha(0)$, we have

$$\alpha(0) = f(\emptyset(0,\overset{o}{x},\overset{o}{v})) = f(\overset{o}{x}) = 0,$$

since $\overset{o}{x}$ is a point on bdry(G). By the chain rule and the fact that $\underline{\overset{o}{v}}$ is a tangent vector to the surface bdry(G) at $\overset{o}{x}$,

$$\dot{\alpha}(0) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} (\overset{o}{x})\overset{o}{v}^i = 0.$$

Since the function $\emptyset$ is a geodesic,

$$\ddot{\alpha}(0) = \sum_{i,j=1}^{n} f_{|ij}(x)\overset{o}{v}^i\overset{o}{v}^j < 0 \ .$$

Thus, for all sufficiently small values of $|s|$, $\alpha(s)$ is negative. In view of our conventions on the function f, it follows that the geodesic $\emptyset(s,\overset{o}{x},\overset{o}{v})$ has no points in common with the open set G for sufficiently small values of $|s|$.

With (1′) it is easy to show that "parallel" surfaces are convex in the same sense as the surface bdry(G) if they are sufficiently close to this surface. By "parallel" surfaces we mean the surface $[\text{bdry}(G)](k)$

where in surface normal coordinates

(Def) $$[\mathrm{bdry}(G)](k) = \{\underline{y} \mid y^1 = k\} .$$

Of course $[\mathrm{bdry}(G)](0)$ is a part of the surface bdry(G). For in surface normal coordinates the equation $f(x) = 0$ becomes $y^1 = 0$ and tangent vectors have the form $(0, y^2, \ldots, y^n)$. Applying the definition of $f_{|ij}$ to inequality (1') of the previous subsection, we have in surface normal coordinates

$$\sum_{i,j=2}^{n} \Gamma^1_{ij}(0, y^2, \ldots, y^n)\overset{\circ}{v}{}^i\overset{\circ}{v}{}^j > 0. \tag{2}$$

For the function $f(\underline{y}) = y^1 - k$ the left-hand side of inequality (1') reduces to

$$\sum_{i,j=2}^{n} \Gamma^1_{ij}(k, y^2, \ldots, y^n)\overset{\circ}{v}{}^i\overset{\circ}{v}{}^j .$$

From the continuity of Γ^1_{ij} and from inequality (2) it follows that for sufficiently small values of $|k|$

$$\sum_{i,j=2}^{n} \Gamma^1_{ij}(k, y^2, \ldots, y^n)\overset{\circ}{v}{}^i\overset{\circ}{v}{}^j > 0$$

for all tangent vectors $(0, \overset{\circ}{v}{}^2, \ldots, \overset{\circ}{v}{}^n)$.

We dignify this assertion by saying that the "parallel" surfaces are <u>convex in the positive sense</u>.

The next lemma relates the concepts of local convexity and convexity of the boundary. Under the general assumptions of this chapter we have the

<u>Lemma</u>. If G has a convex boundary, then the set $\mathrm{cl}(G) = G \cup \mathrm{bdry}(G)$ is locally convex.

Proof. We show a stronger result, namely, that if x and y are sufficiently close points of the boundary of G, then the minimal curve joining these two points lies in G except for the two endpoints x and y.

The proof is indirect. If the contrary were true, there would exist two sequences of points $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$ on the boundary of G such that $x_n y_n \to 0$ as $n \to \infty$ and such that the minimal geodesic connecting x_n to y_n intersects the boundary of G. Since the set bdry(G) is a closed and bounded subset of the finitely compact space cl(G), we may assume without loss of generality that

$$\lim x_n = \lim y_n = z$$

for some point z on the boundary of G.

Let $S(z,\delta)$ be a spherical neighborhood of z in which surface normal coordinates are introduced. Recall that the distance between two points is the intrinsic distance. We see that the minimal curve joining two points of the sphere $S(z,\delta/2)$ lies in the sphere $S(z,\delta)$. Let the integer n be taken so large that x_n and y_n belong to the sphere $S(z,\delta/2)$. For such integers n let $g_n(t)$ be the normal parameterization of a minimal curve joining x_n to y_n (there is only one if the integer n is sufficiently large).

Since (as we assume) $g_n(t_n)$ is a point on the set bdry(G) for some number t_n in the interval (0,1), there is an interval $[a_n,b_n]$, $0 \leqq a_n \leqq b_n \leqq 1$ such that the restriction of g_n to the interval $[a_n,b_n]$, denoted by $g_n \mid [a_n,b_n]$, has no points in common with G. For if the geodesic g_n is tangent to the boundary of G at the point of contact $g_n(t_n)$, the conclusion of our propositions yields the result. If the geodesic g_n is not tangent at t_n, it must penetrate the boundary of G, and we have such an interval of the form $[t_n, b_n]$ or $[a_n, t_n]$.

We may assume without loss of generality that the points $g_n(a_n)$ and $g_n(b_n)$ lie on the boundary of G. Applying our analogue of the theorem of the mean to the geodesic $g_n | [a_n, b_n]$, we obtain a number θ_n in the open interval (a_n, b_n) so that

$$g_n(\theta_n)\mathrm{bdry}(G) \geqq g_n(t)\mathrm{bdry}(G)$$

for all real t in $[a_n, b_n]$. Since the point $g_n(\theta_n)$ is in the exterior of $[G \cup \mathrm{bdry}(G)]$ and since the surface normal coordinates of such points have negative y^1-coordinates, the y^1-coordinate of $g_n(\theta_n)$ is $-g_n(\theta_n)\mathrm{bdry}(G)$. Consider the "parallel" surfaces

$$[\mathrm{bdry}(G)]_n = \left\{ \underline{y} \,|\, y^1 = -g_n(\theta_n)\mathrm{bdry}(G),\ y \in S(\underline{z}, \delta) \right\}.$$

The geodesic g_n is tangent to the "parallel" surface $[\mathrm{bdry}(G)]_n$ at the point $g_n(\theta_n)$ (that is, the geodesic g_n is perpendicular to the y^1-curve at that point by our analogue to the theorem of the mean).

Since

$$\lim g_n(\theta_n) = z$$

and, therefore,

$$\lim g_n(\theta_n)\mathrm{bdry}(G) = 0 ,$$

it follows from the convexity in the positive sense of the surface $[\mathrm{bdry}(G)]_n$ (for integer n sufficiently large and fixed) that the y^1-coordinate of $g_n(t)$ for t near θ_n is less than the number

$$-g_n(\theta_n)\mathrm{bdry}(G).$$

This fact contradicts the maximality of the distance $g_n(t)\mathrm{bdry}(G)$ for $t = \theta_n$.

From our arguments it is evident that the following partial converse is correct.

If the set $\mathrm{cl}(G) = [G \cup \mathrm{bdry}(G)]$ is locally convex and the quadratic form

$$\sum_{i,j=1}^{n} f_{|ij}(\overset{o}{x})\overset{o}{v}^i\overset{o}{v}^j$$

is definite for all tangent vectors $\overset{o}{v}$ to the surface bdry(G) at $\overset{o}{x}$ (i.e.,

$$\sum_{i,j=1}^{n} f_{|ij}(\overset{o}{x})\overset{o}{v}^i\overset{o}{v}^j \neq 0),$$

then

$$-\sum_{i,j=1}^{n} f_{|ij}(\overset{o}{x})\overset{o}{v}^i\,\overset{o}{v}^j > 0 .$$

4.4 The Whittaker Condition

We show in this section that convexity of the set bdry(G) is equivalent to the boundary conditions given by Whittaker for the case considered by him (see Chapter One). In this case our Riemann manifold X is an open subset of the Euclidean plane with fundamental form

$$\lambda_{ij}(x^1,x^2) = [h - V(x^1,x^2)]\delta_{ij}$$

for integers $i,j = 1,2$. The symbols δ_{ij} and x^i denote, respectively, the Kronecker delta and Euclidean coordinates of the point $\underline{x}$ in the plane. The number h is the "energy" of the dynamical system and the function V its potential function, which is defined and has continuous second partial derivatives in X.

We assume that on the manifold X the inequality

$$h - V(x^1,x^2) \geqq \alpha$$

holds for some positive number α. This assumption implies that X is finitely compact. The subset G of X in Whittaker's theorem is a topological image of an open annulus, the boundary of G being two Jordan curves having continuous derivatives through third order. (Actually, continuous second-order derivatives suffice to yield the result in this case.)

In the plane let the function $\underline{\theta}(s)$ be a parametric representation of one of the boundary curves (say the "inner" one) in a neighborhood of the point $\overset{o}{\underline{x}}$. Let the parameter s denote Euclidean arc length so that $\underline{\theta}(0) = \overset{o}{\underline{x}}$. That is, $\sum_{i,j=1}^{2} \delta_{ij}\dot{\theta}^i(s)\dot{\theta}^j(s) = 1$.

Since $\theta(s)$ is a parametric representation of a boundary curve which we assume to be given by the equation

$$f(\underline{x}) = 0,$$

we have

$$f(\underline{\theta}(s)) \equiv 0.$$

Differentiating, we see that

$$0 = \frac{d^2}{ds^2} f(\underline{\theta}(s)) = \sum_{i,j=1}^{2} \frac{\partial^2 f}{\partial x^i \partial x^j}(\underline{\theta}(s)) \frac{d\theta^i}{ds}\frac{d\theta^j}{ds} + \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(\underline{\theta}(s))\frac{d^2\theta^i}{ds^2}. \tag{1}$$

The quadratic form

$$-\sum_{i,j=1}^{2} f_{|ij}(\overset{o}{\underline{x}})\vartheta^i \vartheta^j$$

becomes, by equation (1) above and the definition of $\underline{\theta}(s)$,

$$-\sum_{i,j=1}^{2} f_{|ij}(\underline{\theta}(0))\frac{d\theta^i}{ds}(0)\frac{d\theta^j}{ds}(0) =$$

$$-\sum_{i,j=1}^{2} \frac{\partial^2 f}{\partial x^i \partial x^j}(\underline{\theta}(0))\frac{d\theta^i}{ds}(0)\frac{d\theta^j}{ds}(0) + \sum_{i,j,k=1}^{2} \Gamma^k_{ij}\ (\underline{\theta}(0))\frac{\partial f}{\partial x^k}\ (\underline{\theta}(0))\frac{d\theta^i}{ds}(0)\frac{d\theta^j}{ds}\ (0)$$

$$= \sum_{i=1}^{2} \frac{\partial f}{\partial x^i}\ (\underline{\theta}(0))\frac{d^2\theta^i}{ds^2}\ (0)$$

$$+ \sum_{i,j,k=1}^{2} \Gamma^k_{ij}\ (\theta(0))\frac{\partial f}{\partial x^k}\ (\theta(0))\frac{d\theta^i}{ds}\ (0)\frac{d\theta^j}{ds}(0)\ .$$

Denote by $\underline{\overset{o}{\vartheta}}$ the normal vector whose coordinates are given by

$$\vartheta^i = \sum_{j=1}^{2} \lambda^{ij}(\underline{\overset{o}{x}})\ \frac{\partial f}{\partial x^j}\ (\underline{\overset{o}{x}})\ .$$

Then,

$$-\sum_{i,j=1}^{2} f_{|ij}(\underline{\overset{o}{x}})\overset{o}{\vartheta}{}^i\overset{o}{\vartheta}{}^j = \sum_{i,j=1}^{2} \lambda_{ij}(\underline{\theta}(0))\vartheta^j\ \frac{d^2\theta^i}{ds^2}\ (0)$$

$$+ \sum_{i,j,k=1}^{2} \Gamma_{ij,k}\ (\theta(0))\vartheta^k\ \frac{d\theta^i}{ds}(0)\ \frac{d\theta^j}{ds}(0)\ .$$

From the fact that

$$\sum_{i,k=1}^{2} \overset{o}{\vartheta}{}^i\vartheta^k\ \delta_{ik} = [h - V(\underline{\overset{o}{x}})]^{-1} \sum_{i,k=1}^{2} \overset{o}{\vartheta}{}^i\vartheta^k\ \lambda_{ik}(\underline{\overset{o}{x}}) = 0,$$

we see that

$$\sum_{k,i,j=1}^{2} \Gamma_{ij,k}(\underline{\overset{o}{x}})\overset{o}{\vartheta}{}^i\overset{o}{\vartheta}{}^j\vartheta^k = 1/2 \sum_{k=1}^{2} \frac{\partial V}{\partial x^k}\ \overset{o}{\vartheta}{}^k\ .$$

Hence,

$$-\sum_{i,j=1}^{2} f_{|ij}(\overset{o}{\underline{x}})\overset{o}{\nu}^{i}\overset{o}{\nu}^{j} = \sum_{i,j=1}^{2} [h - V(\overset{o}{\underline{x}})]\delta_{ij}\overset{o}{\nu}^{i}\frac{d^2\theta^j}{ds^2}(0)$$

$$+ 1/2 \sum_{k=1}^{2} \frac{\partial V}{\partial x^k}(\overset{o}{\underline{x}})\overset{o}{\nu}^{k} .$$

Recall that the normal vector $\underline{\nu}$ is selected so that it "points" into the region G. Hence, $\nu^1 = k\cos\gamma$, $\nu^2 = k\sin\gamma$, where

$$k = [h - V(\overset{o}{\underline{x}})]^{-\frac{1}{2}} \left[\sum_{i,j=1}^{2} \lambda^{ij}(\overset{o}{\underline{x}}) \frac{\partial f}{\partial x^i}(\overset{o}{\underline{x}}) \frac{\partial f}{\partial x^j}(\overset{o}{\underline{x}})\right]^{\frac{1}{2}}$$

and γ is the angle specified by Whittaker (i.e., the oriented angle between the x-axis and the outer normal — see Chapter One). On the inner boundary curve

$$\frac{d^2\theta^j}{ds^2}(0) = -\frac{1}{\rho(\overset{o}{\underline{x}})}\left[\delta_{j1}\cos\gamma + \delta_{j2}\sin\gamma\right]$$

for integers $j = 1,2$, where $\frac{1}{\rho(\overset{o}{\underline{x}})}$ is the Euclidean radius of curvature of the inner curve at the point $\overset{o}{\underline{x}}$. The quadratic form becomes

$$-\sum_{i,j=1}^{2} f_{|ij}(\overset{o}{\underline{x}})\overset{o}{\nu}^{i}\overset{o}{\nu}^{j} = k\left[-\frac{h - V(\overset{o}{\underline{x}})}{\rho(\overset{o}{\underline{x}})} + \frac{1}{2}\frac{\partial V}{\partial x^1}(\overset{o}{\underline{x}})\cos\gamma\right.$$

$$\left. + \frac{1}{2}\frac{\partial V}{\partial x^2}(\overset{o}{\underline{x}})\sin\gamma\right]$$

and the hypothesis that this form be positive on the inner curve is seen to be equivalent to Whittaker's condition on the inner curve (k is positive). The equivalence on the outer curve is established in a similar fashion.

SUMMARY

For a Newtonian dynamical system of two degrees of freedom, Whittaker's theorem (Analytical Dynamics, page 389) asserts the existtence of a non-trivial periodic orbit which lies within a ring-shaped region of the x,y-plane. Whittaker's proof is incomplete. By means of Maupertuis' principle we interpret the theorem as an existence theorem for closed geodesic curves in a Riemann manifold. In this study, it is shown that the hypotheses of Whittaker's theorem imply a type of convexity of the ring-shaped region with segments of geodesic curves replacing line segments.

By means of the connection between Riemann geometry and dynamical systems of n-degrees of freedom, Whittaker's theorem is generalized and proved. Our proof is based on Hilbert's approach to the existence of curves of least length and on the convexity properties implied by Whittaker's hypotheses; it is broken into two parts.

The first part of our proof uses an original approach to the intrinsic geometry of distance spaces (metric spaces) which culminates in a proof of the following theorem of Hadamard:

Within each free-homotopy class of a connected, finitely compact intrinsic distance space which has unique line segments in the small, there is a parametric curve whose length is less than or equal to the length of any other curve of this class.

With Hadamard's theorem and with Maupertuis' principle we establish the following generalization of Whittaker's theorem:

Let G denote an open connected subset of the n-dimensional Riemann manifold X of class C^3; let the boundary of G be the set

$$\mathrm{bdry}(G) = \left\{x \mid f(x) = 0\right\}$$

where f is a real-valued function of class C^3 on an open set in X and where (we assume) $f(x) > 0$ for points x which are common to the set G and to the domain of the function f. Suppose that the closure of G,

$$\mathrm{cl}(G) = G \cup \mathrm{bdry}(G),$$

is finitely compact and suppose that for each point $\overset{o}{x}$ in bdry(G)

$$-\sum_{i,j=1}^{n} f_{|ij}(\overset{o}{x})\overset{oi}{v}\overset{oj}{v} > 0 ,$$

where $f_{|ij}$, i, j = 1, ..., n, denote coordinates of the second covariant derivative with respect to the fundamental form of X and where $\overset{oi}{v}$, i = 1, ..., n, denote coordinates of an arbitrary contravariant vector which is tangent to the surface bdry(G) at the point $\overset{o}{x}$, i.e., $\sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(\overset{o}{x})\overset{oi}{v} = 0$. Then in each free-homotopy class of cl(G), there is a curve having minimal length; this curve is a closed geodesic of the manifold X and lies within the open set G.

If the Riemann manifold X has fundamental form

$$[h - V(x)]\lambda(x)$$

where h is a fixed real number, V is a real-valued function of class C^3 and λ is a fundamental form of class C^3, then the closed geodesic of X which lies in G corresponds to a periodic orbit of a Lagrangian dynamical system; the dynamical system has potential V, and the periodic orbit has energy h.

APPENDIX A

CALCULUS OF VARIATIONS

Let f denote a real-valued function with domain the open subset of Euclidean (2n + 1) space

$$\left\{(t,\underline{x},\underline{w}) \mid -\infty < t < \infty, \quad (\underline{x},\underline{w}) \in U\right\},$$

where $\underline{x}$ and $\underline{w}$ denote n-tuples. Suppose that the function f has continuous first partial derivatives; we denote the partial derivative $\frac{\partial f}{\partial x^i}$ by $f_{i|}$ and the partial derivative $\frac{\partial f}{\partial w^i}$ by $f_{|i}$ for each integer i = 1,..., n. A function $\underline{x}$ of the real parameter t will be called <u>admissible</u> if it has the following properties:

1. $\underline{x}$ is defined, continuous and has a continuous derivative on the interval [a,b],
2. $\underline{x}(a) = \overset{0}{\underline{x}}$ and $\underline{x}(b) = \overset{1}{x}$ for fixed n-tuples $\overset{0}{\underline{x}}$ and $\overset{1}{\underline{x}}$, and
3. $(\underline{x}(t), \dot{\underline{x}}(t))$ belongs to the open set U for each time t in the interval [a,b].

<u>Lemma</u> (Euler, Lagrange). If, for the fixed admissible function $\underline{x}$, the inequality

$$\int_a^b f(t,\underline{x}(t),\dot{\underline{x}}(t)) \quad dt \leqq \int_a^b f(t,\underline{z}(t),\dot{\underline{z}}(t)) \quad dt$$

is satisfied for an arbitrary admissible function $\underline{z}$, then the function $\underline{x}$ satisfies the system of ordinary differential equations (Euler's equations)

$$f_{i|} \ (t,\underline{x},\dot{\underline{x}}) - \frac{d}{dt} f_{|i} \ (t,\underline{x},\dot{\underline{x}}) = 0, \ i = 1, \ldots, n.$$

If in addition the function f has continuous second partial derivatives and if the determinant of the matrix with (i,j)-entry $f_{|i|j}$ is non-vanishing for t in [a,b] and $(\underline{x},\underline{w})$ in U, then the function $\underline{x}(t)$ has two continuous derivatives on [a,b].

For the proof of this lemma see [CH; pp. 184-202].

BIBLIOGRAPHY

A. T. M. Apostol, Mathematical Analysis, Addison-Wesley, Massachusetts, 1957.

B. G. D. Birkhoff, Dynamical systems with two degrees of freedom, Trans. Am. Math. Soc. 18 (1917), 199-300.

BR. G. Birkhoff and G. Rota, Ordinary Differential Equations, Blaisdell, Massachusetts, 1969.

Bz. O. Bolza, Vorlesungen über Variationsrechnung, Teubner-Verlag, Leipzig, 1909.

CB. Y. Choquet-Bruhat, Géométrie différentielle et systèmes extérieurs, Dunod, Paris, 1968.

CH. R. Courant and D. Hilbert, Methods of Mathematical Physics, Vol. I, Interscience, New York, 1953.

E. L. P. Eisenhart, Riemannian Geometry, Princeton University Press, New Jersey, 1925.

G. E. Goursat, A Course in Mathematical Analysis, Vol. I, Dover, New York, 1959.

H. J. K. Hale, Ordinary Differential Equations, Wiley-Interscience, New York, 1969.

Hm. P. R. Halmos, Finite-Dimensional Vector Spaces, Van Nostrand Co., New Jersey, 1958.

HWi. P. Hartman and A. Wintner, On the problems of geodesics in the small, Am. J. Math. 73 (1951), 132-148.

HW. W. Hurewicz and H. Wallman, Dimension Theory, Princeton University Press, New Jersey, 1941.

K. E. Kamke, Differentialgleichungen reeller Funktionen, Akademische Verlagsgesellschaft, Leipzig, 1950.

KN. S. Kobayashi and K. Nomizu, Foundations of Differential Geometry, Vol. I, Interscience, New York, 1963.

Kr. E. Kreyszig, Introduction to Differential Geometry and Riemannian Geometry, University of Toronto Press, 1968.

Pj. L. S. Pontrjagin, Topologische Gruppen, Teil 1,2, Teubner-Verlag, Leipzig, 1957.

BIBLIOGRAPHY (Continued)

Py. L. S. Pontryagin, Topological Groups, Second Edition, Gordon and Breach, New York, 1966.

R. W. Rinow, Die inner Geometrie der metrichen Räume, Springer-Verlag, Berlin, 1961.

S. A. Signorini, Esistenza di un'estremale chiusa dentroun contoro di Whittaker, Circolo Matematico Di Palermo (Rendiconti) 33 (1912), 187-193.

St. J. J. Stoker, Differential Geometry, Wiley-Interscience, New York, 1969.

T. H. Tietze, Über Konvexheit im kleinen und im grossen und über gewisse den Punkten einer Menge zugeordnete Dimensionzahlen, Math. Z. 28 (1928), 697-707.

Wd. J. H. C. Whitehead, Convex regions in the geometry of paths, Quarterly J. Math. (Oxford) 3 (1932), 33-52; Addendum, ibid., 4 (1933), 226-227.

Wh. E. T. Whittaker, Analytical Dynamics of Particles and Rigid Bodies, Fourth Edition, Cambridge University Press, London, 1936.

Wi. A. Wintner, The Analytical Foundations of Celestial Mechanics, Princeton University Press, New Jersey, 1941.

VITA

Philip F. Lee, III, was born on March 4, 1942 in New Orleans, Louisiana. He is the younger of two children of Helen Labouisse (formerly Lee) and the late Philip F. Lee, Jr. He was graduated from the Sewanee Military Academy, Sewanee, Tennessee, in 1959. That year he entered Georgia Institute of Technology and was graduated in 1964 with a Bachelor of Science degree in Applied Mathematics.

Mr. Lee immediately began graduate studies at Georgia Institute of Technology, where in 1967, he received a Master of Science degree in Applied Mathematics. During his graduate studies Mr. Lee taught mathematics as a graduate teaching assistant.